UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA BIOMÉDICA

DISEÑO DE UNA METODOLOGÍA PARA EL PROCESAMIENTO DE IMÁGENES MAMOGRÁFICAS BASADA EN TÉCNICAS DE APRENDIZAJE PROFUNDO

ELIA PÉREZ PÉREZ

2017

ETSIT UPM

TRABAJO FIN DE GRADO

Título: **Diseño de una metodología para el procesamiento de imágenes mamográficas basada en técnicas de Aprendizaje Profundo**

Autor: **Elia Pérez Pérez**

Tutor: **Carmen Sánchez Ávila**

Departamento: **Departamento de Matemática aplicada a las Tecnologías de la Información y las Comunicaciones**

TRIBUNAL:

Presidente:

Vocal:

Secretario:

Suplente:

FECHA DE LECTURA:

CALIFICACIÓN:

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA BIOMÉDICA

DISEÑO DE UNA METODOLOGÍA PARA EL PROCESAMIENTO DE IMÁGENES MAMOGRÁFICAS BASADA EN TÉCNICAS DE APRENDIZAJE PROFUNDO

ELIA PÉREZ PÉREZ

2017

RESUMEN

El Aprendizaje Profundo es un subcampo dentro del Aprendizaje de Máquina que utiliza diferentes algoritmos de aprendizaje automático para modelar abstracciones de alto nivel en datos usando arquitecturas jerárquicas, conocidas como redes neuronales profundas (DNNs). Entre los múltiples algoritmos que se pueden encontrar, existen algunos como las redes neuronales convolucionales (CNNs), los autocodificadores y las redes recurrentes (RNNs), que pueden ser de gran ayuda a la hora de analizar imágenes médicas.

El gran potencial que tienen estas técnicas para el análisis de imagen médica reside en su velocidad y eficacia una vez que se tienen una gran cantidad de datos. Su uso se puede aplicar a tareas tan diversas como la detección y segmentación de tumores, así como su seguimiento y control; la visualización y cuantificación del flujo sanguíneo, o a la creación de sistemas de ayuda para interpretación de resultados médicos. Por lo tanto, es lógico pensar que en un futuro serán técnicas cada vez más utilizadas, convirtiéndose muchas de estas tareas algo propio de un ordenador.

Los objetivos de este Trabajo de Fin de Grado son los siguientes:

- La introducción al Aprendizaje Profundo y a los distintos algoritmos que se emplean actualmente, destacando sus ventajas y desventajas.
- Revisar el estado del arte de las técnicas de Aprendizaje Profundo usadas para el análisis de imágenes médicas, así como la identificación de los campos médicos en los que estos algoritmos pueden ser de utilidad.
- La identificación de los algoritmos de Aprendizaje Profundo que pueden emplearse en el análisis de imágenes mamográficas.
- El diseño de una metodología específica para el procesamiento de imágenes mamográficas utilizando las técnicas mencionadas.

Para ello se llevará a cabo un amplio estudio del estado del arte de los diversos algoritmos de Aprendizaje Profundo y de sus usos en el análisis de imagen médica. También se trabajará en la familiarización con algunos de los algoritmos más directamente relacionados con la segmentación de imagen, por su aplicabilidad a la detección de masas y microcalcificaciones en mamografía digital, que serán de vital importancia en la metodología diseñada. Para ello se emplearán distintas fuentes bibliográficas de referencia.

Para finalizar, con este Proyecto se quieren señalar las múltiples aplicaciones que tienen los algoritmos de Aprendizaje Profundo en medicina, y resaltar como su uso ayudará a los médicos a tomar mejores decisiones, así como a mejorar los resultados médicos tanto en términos de tiempo como de eficacia.

PALABRAS CLAVE:

APRENDIZAJE DE MÁQUINA, REDES NEURONALES PROFUNDAS, REDES NEURONALES CONVOLUCIONALES, ANALISIS DE IMAGEN MÉDICA, MAMOGRAFÍA, SEGMENTACIÓN DE IMAGEN.

SUMMARY

Deep Learning is a part of the broader Machine Learning field that uses different automatic learning algorithms for modelling high-level abstractions in data using hierarchical architectures. In general, these architectures are known as deep neural networks (DNN). Among the many algorithms that we can find, there are some such as convolutional neural networks (CNN), autoencoders and recurrent neural networks (RNN), which can be of great help when analysing medical images, especially CNNs.

The great potential of these techniques for medical image analysis lies in its speed and efficiency once a large amount of data is collected. Therefore, it is logical to think that in the future they will be used more and more, becoming this analysis task typical more typical of a computer than of a doctor. Its use is foreseen in fields as diverse as the detection and segmentation of tumours, as well as their monitoring and control; the visualization and quantification of blood flow, or the support decision systems for interpretation of medical results.

The objectives of this Final Degree Thesis are the ones that follow:

- Introducing Deep Learning and the different algorithms that are currently in use, highlighting their advantages and disadvantages.
- Reviewing the state of art of the main DL techniques with a greater performance in medical image analysis, as well as identificating the medical fields in which these algorithms can be useful.
- The identification of the DL algorithms that can be used for analysing mammographies.
- The design of a specific methodology for processing mammographic images using the aforementioned techniques.

For carrying out this thesis a wide study of the State of art of the different Deep Learning algorithms, as well as their use in MIA will be made. Besides, we will also work on familiarization with some of the algorithms more directly related with image segmentation, due to their applicability to mass detection and microcalcifications in digital mammography, being both of vital importance for the methodology that will be designed. For this purpose, different bibliographic sources will be used.

All in all, this Project wants to point out all the potential applications that Deep Learning algorithms have in relation to medicine. Applying these techniques will help doctors to make better decisions and also to improve medical results both in terms of time and efficacy.

KEYWORDS:

DEEP LEARNING, DEEP NEURAL NETWORKS, CONVOLUTIONAL NEURAL NETWORKS, MEDICAL IMAGE ANALYSIS, MAMMOGRAPHY, IMAGE SEGMENTATION.

ÍNDICE

1. Introducción y objetivos

En este capítulo se pretende hacer una introducción a la temática que aborda este trabajo, así como de su propósito y objetivos.

1.1. Introducción

Los avances en Inteligencia Artificial (IA) en los últimos años posibilitan la creación de nuevas tecnologías y tienen cada vez más aplicaciones en diversos campos. Sus fundamentos, complejos algoritmos, pretenden dar solución a todo tipo de problemas intuitivos y subjetivos que hasta ahora, los ordenadores no eran capaces de solucionar, y eran dominio únicamente de los seres humanos, capaces de resolverlos de forma automática (i.e. reconocer a alguien en una foto, o en el campo médico, detectar la presencia de una anomalía en una imagen o en una señal).

Para lograr un buen desempeño por parte de los ordenadores en estos problemas, se busca hacer que puedan aprender de la experiencia, que aprendan la realidad en forma de conceptos simples que se relacionen entre ellos formando otros más complejos, de forma que exista una jerarquía de conceptos, con muchas capas. De este modelo de aprendizaje, con múltiples niveles de representación y abstracción surge lo que se conoce como Aprendizaje Profundo *(Deep Learning).* Los ordenadores que trabajan con algoritmos de Aprendizaje Profundo tienen que ser capaces por si mismos de adquirir su propio conocimiento, extrayendo sus propias deducciones, sus propios patrones de los datos que se le proporcionan. Esta capacidad de aprendizaje es lo que se conoce como Aprendizaje de Máquina *(Machine Learning),* y engloba al ya mencionado Aprendizaje Profundo.

Sin embargo, por muy nuevo que pueda parecer este campo, el Aprendizaje Profundo existe desde hace ya bastante tiempo. La ausencia de fama de estas técnicas se debe a dos motivos principalmente: 1) la no disponibilidad de grandes cantidades de datos para entrenar a los modelos, y 2) no tener ni hardware (HW) ni software (SW) en los ordenadores lo suficientemente potentes como para la ejecución de los modelos. En los últimos años ambos problemas se han visto solventados, y sus múltiples aplicaciones han motivado el creciente interés por este campo. Desde clasificación de imágenes y segmentación de estructuras a reconocimiento de habla o detección de objetos, las posibilidades son muchas. Por ello, hoy en día está siendo usado por compañías tecnológicas, compañías de infraestructuras de SW y para aplicaciones científicas, entre otras.

Las infinitas aplicaciones científicas y médicas de los algoritmos de Aprendizaje Profundo (i.e. para el diagnóstico asistido por ordenador, para hacer predicciones a partir de grandes cantidades de datos, para procesar imágenes en medicina, para diseñar medicamentos o para construir mapas 3D del cerebro) son precisamente la motivación de este trabajo, pues pueden ayudar a mejorar el cuidado de la salud del paciente, en términos tanto de precisión como de rapidez. Una tarea médica en concreto donde estos algoritmos están consiguiendo muy buenos resultados es la de clasificación de lesiones y tumores, es decir, dada una imagen determinada, y basándose en sus características, asignarle a la imagen una de las dos clases de salida posibles (i.e. tumor maligno o benigno). En particular, las investigaciones que los incorporan para la detección temprana del cáncer de mama a partir de mamografías digitales están consiguiendo muy buenos resultados,

ayudando al especialista a tomar mejores decisiones o a fijarse en ciertas partes de la imagen donde pueden estar las anomalías.

Por ello, por ser una aplicación muy extendida, y también por ser el cáncer de mama uno de los más frecuentes a nivel mundial, en este trabajo se quiere poner especial atención en las técnicas de Aprendizaje Profundo que se pueden emplear para el diagnóstico, el seguimiento y la evolución de los tumores en las mamas. Además, tras estudiar el estado del arte en profundidad, se diseñará una metodología basada en un algoritmo de Aprendizaje Profundo cuya finalidad sea la clasificación de dichos tumores.

1.2. Objetivos

El presente trabajo surge con la finalidad, por un lado, de introducir al lector en el campo del Aprendizaje Profundo y explicar los algoritmos más empleados actualmente, así como de estudiar su potencial a la hora de ayudar a un profesional médico en el análisis de imágenes médicas; y por otro, con el objetivo de revisar el estado del arte de los modelos de Aprendizaje Profundo para imagen mamográfica, y a partir de ello diseñar una metodología propia para la detección y clasificación de las lesiones presentes en estas mamografías.

Así, en los primeros capítulos, *"Introducción al Aprendizaje profundo"* y *"Aplicaciones del Aprendizaje Profundo",* se dará prioridad a la actividad investigadora, estudiando los distintos modelos que existen en profundidad y el estado del arte de sus aplicaciones, con especial atención en aquellas para imagen médica.

Esta tarea investigadora también continuará en los capítulos posteriores, pues para diseñar cada una de las fases que componen el sistema de clasificación de lesiones en las mamas es necesario conocer primero el problema, y segundo los trabajos relacionados publicados hasta la fecha. Esto se hará en las dos primeras secciones del capítulo *"Diseño de una metodología para el análisis de mamografías",* y en la tercera sección se explicará con detalle el método diseñado y se comentarán brevemente los resultados esperados, buscando la posibilidad de optimizarlos.

2. Introducción al Aprendizaje Profundo

En este primer capítulo se hace una introducción de forma general al Aprendizaje Profundo. Se discutirán los conceptos principales necesarios para entender el funcionamiento de una red profunda y se explicarán las motivaciones para usarlas, así como algunas de sus ventajas. También se introducirán los principales tipos de algoritmos de Aprendizaje Profundo, y se comentarán sus ventajas y desventajas. Todo esto es imprescindible para entender más adelante el sistema descrito en el Capítulo 4. Para ampliar esta información se recomienda leer los trabajos [1][2][3][5], a partir de los cuales se han elaborado este Capítulo 2.

2.1. Conceptos básicos

Como ya se introdujo en el sección anterior, tanto el **Aprendizaje Profundo** como el **Aprendizaje de Máquina** son dos formas de **Inteligencia Artificial**. Los métodos del primero se pueden englobar dentro del segundo a su vez, y por ello

para comprender en qué consiste el Aprendizaje Profundo es esencial entender primero las bases en las que está cimentado el Aprendizaje de Máquina.

Para empezar, se debe tener claro cual es el objetivo del **aprendizaje de máquina** (o aprendizaje automático). Podríamos decir que es el desarrollo de sistemas que pueden cambiar su comportamiento de manera autónoma basados en su experiencia y, por tanto, puede aprender a partir de unos datos dados. Aprender hace referencia a que el algoritmo, a partir de la experiencia, puede mejorar su desempeño, ser más hábil, a la hora de realizar unas determinadas tareas. Estas **tareas** son las aplicaciones que se le van a dar al algoritmo; la clasificación de las entradas en distintas categorías, la regresión o predicción de un valor dado una entrada determinada, la transcripción de una entrada no del todo estructurada en una salida de texto discreta, la traducción de máquina, la detección de anomalías, la síntesis y muestreo para la generación de más datos, la predicción de valores dadas entradas vacías, la limpieza de datos, o la estimación de densidad y la estimación de la función de probabilidad de masas. Estas tareas son solo ejemplos, y algunas de ellas serán ampliadas; las relevantes para visión por ordenador y para el análisis de imágenes médicas, junto con otras más adelante en el Capítulo 3.

Para medir el desempeño del algoritmo en una cierta tarea se emplean distintas medidas, de las cuales la más común, sobre todo para tareas de clasificación, es la **precisión** del modelo, que hace referencia a la proporción del modelo que predice la salida de forma correcta. Esta información es equivalente a la proporcionada por la **tasa de error**, la proporción de salidas predichas incorrectamente. Para evaluar el desempeño del algoritmo correctamente lo adecuado es emplear un conjunto de datos distinto al usado para entrenarlo. Aquí aparecen los conceptos de conjunto de datos de entrenamiento y de prueba, que se explicarán más adelante.

Los algoritmos de aprendizaje de máquina pueden clasificarse, de forma general, en supervisados y no supervisados. Esta clasificación hace referencia a los datos con los que experimentan durante su entrenamiento. Un algoritmo es de **aprendizaje no supervisado** cuando trabaja con un conjunto de datos con muchas características del cual aprende propiedades para estructurar esos datos. En el contexto de Aprendizaje Profundo estos algoritmos suelen tener que aprender la distribución de probabilidad del conjunto de datos o ciertas características acerca de este conjunto, pero en cualquier caso tiene que hacerlo por sí mismo, sin tener ningún tipo de guía o de ayuda. Por el contrario, los algoritmos de **aprendizaje supervisado** emplean un conjunto de datos formado por ejemplos, o lo que es lo mismo, **instancias**, asociados con una etiqueta. Así, en este caso, el algoritmo aprende a clasificar las instancias basándose en estas etiquetas, como si tuviera un profesor que le enseñara al sistema qué es lo que tiene que hacer. Aunque la mayoría de algoritmos suelen trabajar de una de estas maneras, existen más variantes que hacen referencia a este paradigma de aprendizaje, como son el aprendizaje semi-supervisado, el aprendizaje multi-instancia, y aprendizaje reforzado (*reinforcement learning*).

Para describir el conjunto de datos se tienen las **características**, cuya elección es esencial para lograr un buen desempeño del algoritmo. Lo más común es agruparlas en una matriz de diseño, siendo cada instancia un vector, y siendo todos estos vectores del mismo tamaño (i.e. para un conjunto de datos formado por

fotografías, todas ellas en un principio tendrían que ser del mismo tamaño, lo cual es difícil. Este punto se analizará más adelante cuando se trate la parte de análisis de imágenes médicas). Para la representación de características existe otra parte de métodos dentro del Aprendizaje de Máquina, que a su vez agrupan a los basados en Aprendizaje Profundo, que se conocen como métodos de **Aprendizaje de Representación** o *representation learning,* que exploran técnicas para averiguar características útiles de los datos a partir de toda la información proporcionada.

De este conjunto de datos inicial no se emplean todos para el mismo propósito, si no que se suelen particionar. Para que el algoritmo aprenda se suele usar el porcentaje mayor de los datos, lo que llamamos **conjunto de datos de entrenamiento**. Una vez entrenado el modelo, se ejecuta sobre un conjunto de datos diferente, el **conjunto de prueba,** para ver cómo generaliza. Este poder de **generalización** es un indicador de lo bien que funciona el algoritmo, y se puede ver midiendo tanto el error de entrenamiento como el error de prueba (o error e generalización), y siempre se busca el disminuirlos al máximo. Como su propio nombre indica, el error de entrenamiento se mide en el conjunto de entrenamiento, y el error de prueba se mide en el conjunto de prueba. Para hacer la susodicha división de los datos, se suele seguir una estrategia de generación de datos, en la cual se asume que los datos de cada conjunto son independientes y que están distribuidos de forma idéntica.

Para lograr la situación ideal de generalización óptima (que el modelo se ajuste bien a nuevos datos), tiene que cumplirse que tanto el error de entrenamiento como la diferencia entre el error de entrenamiento y el error de prueba sean pequeños. Pero esto no siempre se cumple, dando lugar a lo que se conoce como ***underfitting***, cuando el modelo es demasiado sencillo como para captar la complejidad de los datos (se asocia con un error de entrenamiento más alto de lo adecuado) y ***overfitting***, cuando el modelo queda demasiado ajustado a las características de los datos con los que ha sido entrenado y no puede generalizarse (se relaciona con una diferencia entre el error de prueba y el de entrenamiento demasiado grande). Estos dos conceptos están estrechamente relacionados con el de **capacidad**, que es la habilidad del modelo a adaptarse a una amplia variedad de funciones. Si la capacidad del modelo es baja, éste no se adaptará bien a los datos de entrenamiento, y si es alta se adaptará demasiado. Hay que buscar siempre la forma de lograr una capacidad intermedia entre ambas situaciones.

Siguiendo en la línea de entender el comportamiento del algoritmo existen una serie de parámetros denominados **hiperparámetros** con los que se controla la capacidad del algoritmo, entre otras cosas. El problema es que muchos de estos hiperparámetros no pueden ser aprendidos en los datos de entrenamiento, pues nos llevarían de nuevo al problema de *overfitting* y por lo tanto a resultados erróneos. Para solucionar esto aparece un tercer conjunto de datos, el **conjunto de validación**, formado por parte de los datos del conjunto de entrenamiento, y que se usa para estimar el error de generalización durante y tras el entrenamiento. Esto resulta útil para poder ajustar sobre la marcha los hiperparámetros.

En cuanto a las estrategias que se siguen para la división de los datos en entrenamiento y validación, lo más común es usar porcentajes, cumpliendo un ratio 80:20, respectivamente. Otro tipo de táctica habitual es la *cross-validation* o

validación cruzada (CV), en cualquiera de sus diferentes formas (i.e. validación cruzada dejar-uno-fuera o *leave-one-out*, validación cruzada k-veces, etc.). La CV consiste básicamente en repetir las fases de entrenamiento y de validación en distintos subconjuntos de datos escogidos de forma aleatoria.

Otros conceptos de interés a la hora de calificar el desempeño de un algoritmo de aprendizaje de máquina son aquellos relacionados con la estadística, como el sesgo, la varianza, y el error estándar, que estiman el funcionamiento del algoritmo. El **sesgo** o *bias* es el error asociado a asunciones incorrectas a la hora de entrenar, y está relacionado con el *underfittin*g, mientras que la **varianza** se asocia al *overfitting*, y hace referencia a una alta sensibilidad a pequeñas variaciones en los datos.

Hasta ahora, todos estos conceptos son conceptos compartidos por algoritmos de Aprendizaje de Máquina y de Aprendizaje Profundo, pero cuando se trabajan con modelos de este segundo grupo, aparecen nuevos términos que es conveniente explicar.

Se parte del hecho de que un modelo de Aprendizaje Profundo está formado por una serie de **capas**, a su vez formada por **unidades o neuronas**, y que mediante la adición de más capas y/o neuronas se logra que una **red** represente funciones y características de complejidad creciente.

El término **red** hace referencia a que el modelo se estructura agrupando múltiples funciones, a modo de cadena. La primera función es la que se conoce como primera capa de la red, la segunda como segunda capa, tercera capa, cuarta capa, y así sucesivamente. Esta primera capa también se conoce como **capa de entrada**, pues es a la que le llegan los datos, del mismo modo que a la última capa de la red se le denomina **capa de salida.** El resto de las capas, entre la de entrada y la de salida se denominan **capas ocultas,** y de ellas no se muestra nunca su salida. La longitud de esta cadena, el número de funciones que se tengan, es la **profundidad de la red.**

El término **neurona** para denominar a cada una de las unidades que componen una capa, así como el hecho de llamar a los modelos "neuronales" se debe a que en un principio estos modelos se inspiraron en la Neurociencia (si bien hoy en día esta asociación está más en desuso porque la finalidad de estas redes no es modelar el cerebro ni mucho menos, si no el conseguir unos buenos resultados que puedan ser generalizados dado un problema en concreto). Las unidades o neuronas actúan en paralelo, y determinan la anchura del modelo. Cada unidad recibe varias entradas de otras unidades y calcula su propio valor de activación.

Las capas formadas por neuronas tienen que conectarse entre ellas, lo cual se define a partir de la **matriz de pesos.** Por ejemplo, si la matriz de pesos define una transformación lineal, toda unidad de entrada estará conectada a toda unidad de salida, lo que implica múltiples conexiones. El tener muchas conexiones supone a veces un problema, en términos computacionales, y por ello también es habitual aplicar algún tipo de estrategia a la red para reducir las conexiones.

Así, el número de capas de red, el número de unidades de cada capa, cómo se conectan las capas entre ellas y otras decisiones de diseño se engloban en el concepto de **arquitectura de la red.**

En otras palabras, escoger la arquitectura de la red consiste en decidir acerca de la anchura de sus capas y de su profundidad. Aunque las redes neuronales se denominen "profundas" es importante no equivocarse con el significado de este concepto, pues no es necesario tener muchas capas para que nuestra red resulte eficiente; una sola capa oculta, además de la de entrada y de la de salida suelen ser suficientes, y mejor opción que muchas, para adaptarse al conjunto de entrenamiento y poder generalizarse. Si se hacen las redes más profundas, añadiendo capas, se compensa usando menos neuronas por cada capa y menos parámetros, pero esto implica también que sean más difíciles de optimizar. Así, como no hay un modelo preestablecido para elegir la arquitectura, lo mejor es siempre experimentar con ella y ver los resultados que se obtienen con las diferentes opciones propuestas, teniendo como guía al conjunto de validación. A esta conclusión se puede llegar también si se piensa en la red como una estructura a capas, en la que cada capa es un paso que hace uso del resultado obtenido en el paso anterior. Así, en el primer paso el modelo aprende una característica simple, en el segundo una un poco más compleja a partir de la anterior, y así sucesivamente. Esta es una comparación común y fácil para comprender cómo funciona un algoritmo de una red neuronal profunda.

En las redes neuronales, el concepto de entrenamiento, del cual ya se ha hablado, también juega un importante papel. El entrenamiento requiere siempre distintas y múltiples decisiones de diseño (i.e. elegir el optimizador, la función de coste, la forma de las unidades de salida, etc.). En concreto, para las capas ocultas es importante la elección de la **función de activación**, que es la que se encarga de calcular un nuevo valor de salida a partir de todos los valores de entrada que le llegan. Lo habitual es escoger funciones de activación no lineales (i.e. la función sigmoide), pues las lineales solo serían válidas para resolver problemas muy sencillos.

El modelo más representativo de Aprendizaje Profundo es red neuronal con alimentación positiva (También conocida como *Deep Feedforward Network, Feedforward Neural Network, multilayer perceptrons*) o lo que es lo mismo, un **perceptrón multicapa** (MLP), base de muchos otros modelos detallados en las siguientes páginas. En la Figura 1 se muestra la estructura de un MLP con tres capas ocultas entre entrada y salida, para la clasificación de imágenes. Su funcionamiento comienza por introducir en la capa de entrada los píxeles, datos que por sí mismo un ordenador no podría interpretar llegando a un resultado de clasificación coherente. En lugar de intentar hacer un mapeo píxel-clase de forma directa, lo hace por pasos, de forma más simple, asignando a cada capa una característica distinta. Así, comienza por mapear los píxeles a las esquinas, las esquinas a los contornos y bordes, éstos a las partes de los objetos, y finalmente, se identifica el objeto en cuestión.

Si uno sigue la explicación del funcionamiento de este MLP, se puede entender claramente el concepto de **alimentación positiva** (o de propagación hacia delante, *feedforward*) **,** que hace referencia a que la información se empieza a evaluar en la entrada y termina en la salida sin volver hacia atrás. Así, la información es recogida de todas las entradas por una función de agregación, y la salida es calculada a partir de esa misma función y evaluada con una función de activación. Todos estos cálculos necesarios para definir la salida tienen lugar en las capas ocultas de la red, entre la entrada y la salida. El concepto opuesto a este sería

el de **retroalimentación,** que es el fundamento de las Redes Neuronales Recurrentes (RNNs). La retroalimentación, también conocida como retropropagación o *back-propagation* es un algoritmo de entrenamiento de las redes en el cual la información puede fluir hacia atrás, de forma que calcula el gradiente de las funciones que constituyen la red neuronal profunda.

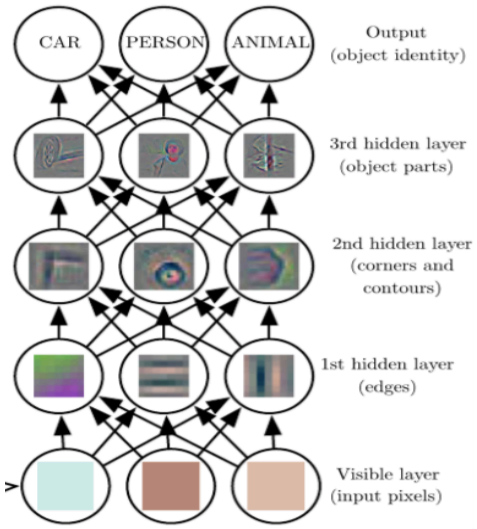


Figura 1. Perceptrón multicapa con tres capas ocultas para la clasificación de imágenes. Obtenida de [1].

Desde una perspectiva un poco más compleja, si nos adentramos en los elementos propios del diseño de la red, para empezar es importante hacer una elección adecuada de la **función de coste.** Una opción frecuente es emplear la máxima verosimilitud (*maximum likelihood*), lo cual significa usar la entropía cruzada entre los datos de entrenamiento y las predicciones del modelo como función de coste; con una forma específica según el modelo en el que estemos trabajando. Una de las grandes ventajas que presenta es que evita el diseño propio de funciones de coste para cada modelo en particular. La salida de esta función son siempre distribuciones de probabilidad, pero también se puede buscar otro tipo de salidas, como medidas estadísticas.

La elección de la función de coste está ligada a la elección de la unidad de salida, que de forma habitual es o bien lineal, o sigmoide o *softmax* [12], que se emplean, respectivamente, para hacer la media de una distribución gaussiana condicional, para predecir el valor de una variable binaria, y para representar la distribución de probabilidad en las *n* clases diferentes, a modo de clasificador. Aunque estas son las tres formas más comunes, se le puede dar prácticamente cualquier forma deseada a esta capa de salida.

El diseño de las capas ocultas de la red también es algo importante a la hora de construir un modelo, si bien todavía no existen una serie de principios teóricos claros sobre cómo hacerlo. Entre los muchos tipos de unidades ocultas, se pueden destacar las unidades lineales rectificadas, ya que suelen ser una buena elección en la mayoría de los casos. Otras opciones son unidades ocultas lineales, ocultas *softmax*, ocultas *softplus*, ocultas RBF (*Radial Basis Function*), etc.

Para terminar con este apartado se introducen dos de conceptos en relación con la **regularización,** definida como "cualquier modificación que se hace a un algoritmo de aprendizaje con la intención de reducir su error de generalización

pero no su error de entrenamiento" [1]. Para lograr esta finalidad se pueden o poner restricciones al modelo o usar métodos de conjunto *(ensemble methods).* De estas últimas destaca el **aumento de datos,** muy usado en problemas de clasificación, y en particular para el reconocimiento de objetos, tanto por su sencillez como por sus buenos resultados (i.e. operaciones de traslación, convolución, de escalado, etc., pero siempre con cuidado de no modificar las clases de salida). Otras estrategias comunes son la adición de ruido a las entradas, como en el caso del autocodificador *denoising*, y la **compartición de parámetros o *parameter sharing***, que fuerza a determinados grupos de parámetros a ser iguales, siendo algo ventajoso en términos de memoria. Esta técnica es muy usada en las redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs) para visión por ordenador, pues permite incrementar de forma significativa el tamaño de las redes sin la necesidad del correspondiente aumento de los datos de entrenamiento. En la siguiente sección de describirán en detalle las características de las CNNs.

2.2. Principales métodos de Aprendizaje Profundo

Con el creciente interés por este campo, cada vez existen más tipos de algoritmos de Aprendizaje Profundo, que van surgiendo como modificaciones de otros para objetivos específicos y concretos.

No es posible hacer una clasificación de manera estricta y cerrada, pues en función de los autores y del enfoque, estas clasificaciones varían. Por ello, se ha escogido la clasificación que ha resultado más adecuada para este trabajo, de acuerdo con la presentada en [3]. Los algoritmos revisados se agrupan así en las cuatro categorías siguientes: 1) Redes Neuronales Convolucionales, 2) Máquinas de Bolztmann restringidas, 3) Autocodificadores y 4) Codificación dispersa; que se plasman de forma más aclaratoria, junto con sus variantes, en la Figura 2. De estas cuatro categorías se explicarán las nociones principales y se analizarán tanto sus contribuciones como sus limitaciones.

Por otro lado, se prestará especial atención en explicar las CNN [6], [7] por ser las que más aplicaciones tienen en análisis de imágenes médicas y en particular en el análisis de mamografías, como se verá en los Capítulos 3 y 4.

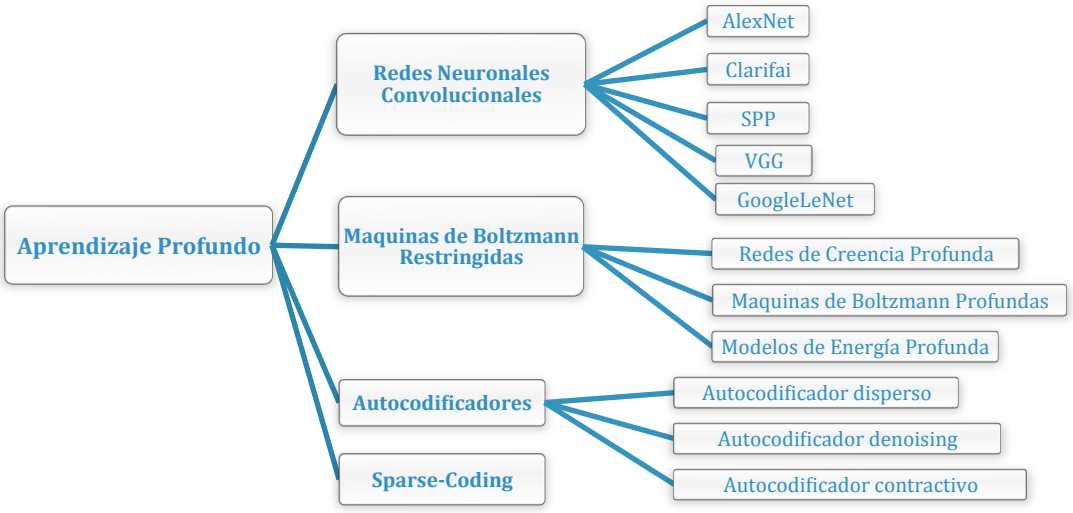


Figura 2: Esquema de agrupación seguido para los distintos métodos de Aprendizaje Profundo.

2.2.1. Redes neuronales convolucionales (CNNs)

Las redes neuronales convolucionales o CNNs, de su nombre en inglés, son tal vez la parte de la IA que más se ha visto inspirada por la Biología, en concreto por la Neurociencia, ya que su funcionan de forma análoga a como lo haría el córtex visual primario (i.e. En ambos casos se define la información almacenada en mapas de características 2D, y las CNNs están formadas por unidades básicas individuales agrupadas en capas de agrupamiento, emulando a las células del córtex que se agrupan en forma de células complejas).

Las CNNs representan un modelo profundo muy exitoso, y hoy en día son empleadas en gran cantidad de aplicaciones, obteniendo por lo general muy buenos resultados. En particular son muy usadas en tareas de visión por ordenador, y por ello también en análisis de imagen médica, por lo que son de gran interés para este trabajo. Está constatado que para este tipo de imágenes son las redes más robustas y exitosas que hay por ahora, como se demostró en el 2012, cuando una CNN ganó el desafío de reconocimiento de objetos de *ImageNet* [8], momento a partir del cual solo se han logrado mejoras [3].

A la hora de definir una CNN se puede decir que es el tipo de red neuronal para el procesado de datos con una topología conocida y cuadriculada, como por ejemplo los datos de una imagen que conforman una cuadrícula 2D formada por píxeles. En esta definición está intrínseca una gran ventaja, el hecho de que las entradas puedan ser de diferentes tamaños, algo en especial muy útil cuando se trabaja con imágenes.

En cuanto a su estructura, las CNN están formadas por tres tipos de capas diferentes; las de convolución (CONV), las de agrupamiento y las totalmente conectadas (*fully-connected layers,* FC*).* Un esquema muy típico es el que se puede ver en la Figura 3, donde la red está formada por 5 capas CONV con varias capas de agrupamiento y seguidas de 3 capas FC, aunque estos números pueden cambiar [7]. A continuación se explica con detalle cada una de estas capas.

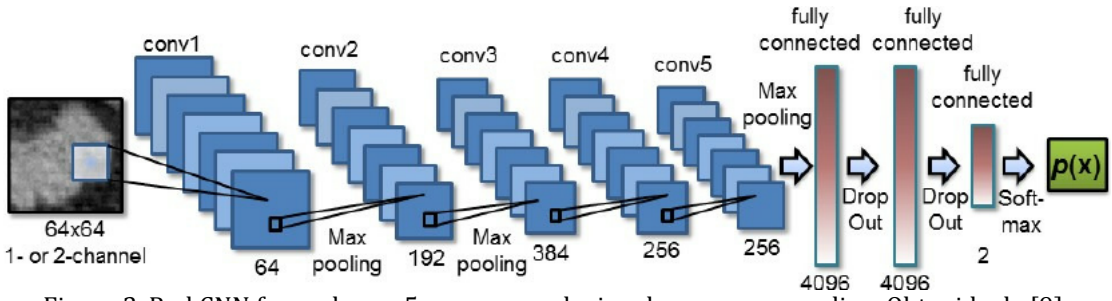


Figura 3. Red CNN formada por 5 capas convolucionales con max_pooling. Obtenida de [9].

1) Capas convolucionales (CONV)

La operación de convolución es la que da nombre a la arquitectura y se emplea en al menos una de sus capas, tanto a la imagen completa como a los mapas de características intermedias, generando nuevos mapas de características. Tiene tres principales ventajas que hacen que el sistema de Aprendizaje Profundo resulte más eficiente:

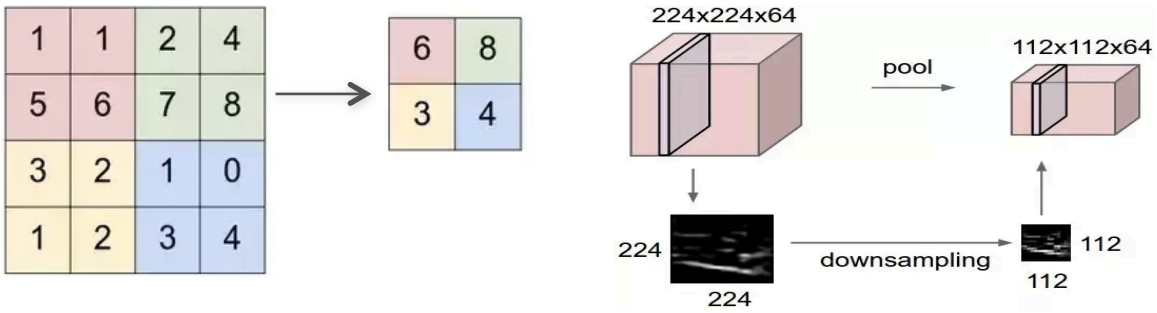
a) Las conectividad o interacciones dispersas. En lugar de haber un parámetro por cada interacción entre cada entrada y cada salida, estos parámetros se comparten, con lo que hay menos parámetros que interacciones. De este modo los requisitos de memoria son menores, además de que se aprenden correlaciones entre píxeles vecinos.
b) La compartición de parámetros. Está ligada al concepto anterior; varias funciones del modelo emplean los mismos parámetros. Se ahorra memoria.
c) La invarianza a la localización del objeto. Deriva de la equivarianza, que hace referencia a que si la entrada cambia, la salida también lo hace en la misma medida.

Debido a estas ventajas en algunas ocasiones se reemplazan las capas FC por capas CONV, para acelerar el proceso de aprendizaje. Esto sucede por ejemplo en las técnicas NIN *(Network to Network)* [10].

2) Capas de agrupamiento o de *pooling*

Las capas de agrupamiento son de gran utilidad pues proporcionan la localización de las características en las imágenes cuando no es necesario conocer los píxeles correspondientes a estas características con exactitud.

Generalmente se sitúan tras cada capa CONV con el objetivo de reducir las dimensiones de los mapas de características. Para ello, aplican una función a las salidas más próximas a la salida de la red, modificando esta última. Se pueden emplear distintos de funciones (*average_pooling, norm_pooling, etc),* pero la más utilizada con diferencia es la de *max_pooling* [3], que reduce la dimensionalidad de la entrada tomando para cada conjunto rectangular de tamaño fijo y escogido, el valor máximo de los píxeles de esa región, y haciendo de ese valor el nuevo valor del píxel de salida. Esto se puede hacer en 1D, como en el caso de la Figura 4 (a), donde un mapa de características de 4x4 queda reducido a 2x2 al aplicar un *max_pooling* de 2x2; o en 2D, muy útil para el submuestreo de imágenes y de *patches*, representado en la Figura 4 (b), y donde se puede ver que el número de capas (64) permanece intacto.



(a) (b)

Figura 4. Funcionamiento del operador de *max-pooling* en un mapa de características 1D (a) y 2D (b). Imágenes obtenidas de [11]

Por sus múltiples usos en imágenes y sus ventajas de reducción de dimensiones e invarianza, las capas de agrupación son las más estudiadas de los tres tipos de capas. De este intenso estudio nacen tres distintos enfoques, cada uno

con distintos propósitos y para distintos procedimientos, aunque se defiende que la mejor práctica es combinar los tres para lograr un buen desempeño de la CNN [3].

a) *Pooling* estocástico. Equivalente al *max pooling,* pero se generan muchas copias de la imagen de entrada, cada una de ellas con pequeñas deformaciones locales. Su naturaleza estocástica, el escoger aleatoriamente la activación dentro de cada región de neuronas, hace que se solucione el problema del *overfitting* que muchas veces provoca el *max_pooling*.

b) Agrupamiento espacial de pirámides (*Spatial Pyramid Pooling*, SPP*)*. En este caso la última capa de agrupamiento se reemplaza por una SPP, capaz de extraer representaciones de longitud fija de imágenes, de manera que se omite la limitación de requerir una imagen de entrada de tamaño fijo. Es una estrategia que se aplica de forma general a las CNN para incrementar su rendimiento.

c) *Def-pooling*. Es un tipo de capa que se introduce en algún punto de la red para poder trabajar con las deformaciones de los objetos de forma más eficiente, lo cual es un desafío en el campo de visión por ordenador, y particularmente en el reconocimiento de objetos.

3) Capas totalmente conectadas (FC)

Estas tipo de capas se sitúan tras la última de las capas de agrupación para convertir los mapas de características 2D en un vector 1D, mucho más útil para la representación de estas características más adelante. Las capas FC trabajan como una red neuronal tradicional y contienen aproximadamente el 90% de los parámetros de la red [7].

El vector 1D de salida de la red suele ser de longitud predefinida, por ejemplo para una tarea de clasificación de imágenes será de longitud igual al número de categorías que se tengan. Otra opción común en el análisis de imágenes es tomarlo como un vector de características 1D para un procesado posterior (lo que más adelante se llamará extracción de características).

La principal desventaja de estas capas es la gran carga computacional que generan a la hora de entrenarlas, debido a la cantidad de parámetros que manejan. Por ello muchos autores defienden disminuir las conexiones entre las neuronas de estas capas empleando algún tipo de método, como en el caso de *GoogleLeNet* [13] , reducirlas en número, o incluso eliminarlas [3].

Finalmente se explica el entrenamiento de las CNNs, pues es la parte más complicada cuando se trabaja con este tipo de arquitecturas. Si se decide emplear una estrategia supervisada, al tener que hacerse la propagación por toda la red hacia delante y hacia atrás, se necesitan ordenadores muy potentes. Así, se sugiere emplear alternativas como pueden ser *Dropout [14][15]* y *DropConnect [16]*, en las que el algoritmo omite la mitad de los detectores de características para que el modelo generalice mejor, inicializar la red con parámetros pre-entrenados en lugar de aleatorios, ajustar los parámetros de la red a la tarea para la cual se quiere emplear, y finalmente aplicar técnicas de aumento de datos, muy empleadas a la hora de trabajar con imágenes médicas pues alivian la necesidad de tener grandes cantidades de datos etiquetados. Del aumento de datos se hablará en el Capítulo 4, si bien cabe mencionar en este punto que gracias a ellas se puede realizar un

entrenamiento no supervisado, y combinarlo con el supervisado. No obstante, existen más estrategias que se detallarán más adelante.

Por los múltiples usos que se les da a las CNNs en el dominio de la visión por ordenador, existen distintos modelos de CNNs muy populares, de los que se resaltan los siguientes:

AlexNet es una de las arquitecturas CNN más conocidas, que sigue la estructura típica de 5 capas CONV y 3 FC. Este modelo fue el responsable del auge de las CNNs cuando se entrenó en *ImageNet* obteniendo resultados remarcables [17], y hoy en día se sigue empleando para la tarea de clasificación de imágenes. En ella se pueden encontrar distintas técnicas de aumento de datos basadas en transformaciones geométricas. Aun así, presenta el inconveniente de que requiere una resolución fija de imagen de entrada (224x224, en concreto), además de que todavía no existe una clara comprensión de por qué funciona tan bien.

Clarifai es otro modelo que surgió al intentar dar explicación a cómo funcionaban internamente las capas intermedias de las arquitecturas CNN. Su estructura es la convencional y también logra muy buenos resultados, mejores que los *AlexNet* en la clasificación de imágenes, como para *ImageNet.*

SPP o Red con Agrupamiento Espacial de Pirámides es otro modelo cuya motivación es evitar el requisito de una resolución fija de imágenes de entrada, para lo cual usa SPPs. Al introducirse las capas SPP en distintos modelos de CNNs [18] se ha logrado aumentar la precisión de los mismos, por lo que esta estrategia de agrupación ha logrado una gran popularidad.

VGG es una CNN profunda empleada para el reconocimiento de imágenes, que en lugar de emplear 5 usa entre 13 y 15 capas CONV, de ahí su profundidad. VGG obtiene muy buenos resultados a la hora de clasificación de imágenes, demostrando así que el incrementar la profundidad de la red puede ir unido a aumentar su precisión. En el caso de VGG esto es posible porque emplea filtros convolucionales muy pequeños en todas las capas. Otra ventaja que aporta es una mayor capacidad de generalización para otros conjuntos de datos [19].

GoogLeNet es otra CNN todavía más profunda que VGG. Está formada por 22 capas CONV, y por una única capa FC. Al emplear esta red en tareas de clasificación de imágenes también se observa un aumento en la precisión de los resultados.

Aunque los resultados obtenidos por estas CNNs en cuanto a la clasificación de imágenes son muy satisfactorios, en estos dos últimos años han surgido otras arquitecturas profundas que buscan mejorarlos, como son *Inception* [20] [21], *ResNet* [22], *y DenseNet* [23]. Estos tres modelos resultan particularmente interesantes para este trabajo, pues la arquitectura que se pretende diseñar está fundamentada en ellos. Por ello, se detallará la estructura de estas redes en el Capítulo 4.

Finalmente, comentar otros dos tipos de redes que derivan de las anteriores, y que tienen otras aplicaciones, en concreto la detección de objetos y la segmentación semántica. Las ***RCNN*** (regiones con características CNN) son básicamente una combinación de CNNs con un SVM lineal (*Support Vector Machine*) empleadas para detectar objetos. Comienzan por generar múltiples propuestas de objetos o candidatos, de los cuales extraen sus características usando la CNN, y finalmente clasifican cada candidato con el SVM en categorías. Esta estrategia de

reconocimiento por regiones consigue muy buenos resultados, y por ello es ampliamente usada en visión por ordenador y también en imagen médica. Sin embargo, su rendimiento está limitado tanto por el hecho de necesitar la red la ubicación del objeto en la imagen como por la gran cantidad de candidatos que genera. A diferencia de las RCNNs, las **FCNs** (red completamente convolucional) se utilizan principalmente para la segmentación semántica, y son útiles para eliminar la restricción de la resolución de la imagen.

2.2.2. Máquinas de Boltzmann restringidas (RBMs)

Las RBMs son un tipo de redes neuronales de estocástica generativa. Son modelos basados en "energía" (*energy-based models*), es decir, modelos con una función de energía compuesta por varios términos, donde cada termino se corresponde a un factor en la distribución de probabilidad. Cada uno de estos términos puede ser entendido como un "experto" que determina si una restricción en concreto se satisface. Son, por tanto, modelos probabilísticos en los que la salida se expresa en función de probabilidades, lo cual es ventajoso para la interpretación humana.

Hoy en día el término RBM se emplea de forma generalizada, para denominar a cualquier modelo con variables latentes, las cuales se agrupan en una sola capa e interaccionan con el resto de capas parametrizadas por matrices, aprendiendo de esta manera la representación de la entrada. Se pueden entender más bien como modelos gráficos usados para componer y entrenar otros modelos de aprendizaje profundo, y no como tales, aunque comparten diversas características con ellos (i.e. sus unidades se organizan en capas, la conectividad entre capas se describe por medio de una matriz, la conectividad es relativamente densa, etc.).

En cuanto a las aplicaciones de las RBMs destacan la segmentación facial y el reconocimiento telefónico, si bien sus usos en imagen médica y en general en visión por ordenador están mucho más reducidos que en el caso de las CNNs.

Todos los RBMs pertenecen al conjunto de Máquinas de Boltzmann, con la principal modificación de que exigen que las unidades visibles estén en igual proporción que las ocultas, formando así un gráfico bipartito. Esta división se impone para hacer a los algoritmos más eficientes. Las principales variantes de las RBMs a considerar, las que más se emplean en tareas de visión por ordenador, son las tres que siguen, cuya principal diferencia es el tipo de conexiones que hay entre las unidades que forman las distintas capas.

1) Redes de creencia profunda (*Deep Belief Networks*, DBNs)

Las DBNs fueron uno de los primeros modelos no convolucionales que mostraron un buen funcionamiento a la hora de entrenar arquitecturas profundas. Desde ese momento, se ha investigado cómo trabajar con ellas y mejorar sus resultados, pero las constantes dificultades encontradas en su etapa de entrenamiento han hecho que su uso haya caído con el paso de los años.

La arquitectura de las DBMs se caracteriza por estar formada por varias capas con variables latentes, que son las capas ocultas, y típicamente binarias, además de la capa de entrada y la de salida, que pueden ser binarias o reales. Estas capas están conectadas solo entre neuronas de capas vecinas, pero nunca entre unidades de la misma capa, y son no dirigidas entre las dos primeras capas y

dirigidas hacia la capa más cercana a los datos en el caso de todas las demás. La salida de estas redes es siempre en forma de distribución de probabilidad conjunta sobre datos y etiquetas observables.

Para entrenarlas primero hay que inicializar capa a capa la red, y luego ajustar todos los pesos de forma conjunta con las salidas deseadas. Este procedimiento no supervisado presenta dos ventajas; primero que la red siempre se inicializa de una manera adecuada, y segundo, que no se requieren datos etiquetados para entrenar. Por el contrario, tiene la desventaja de implicar un alto coste computacional. Además, si se piensa en las tareas relacionadas con visión por ordenador, también hay que considerar el problema de que no tienen en cuenta la estructura 2D de la imagen de entrada, y por estos dos motivos es por lo que su uso es mínimo en los modelos más recientes.

2) Máquinas Profundas Boltzmann (DBMs)

Los DBMs, a diferencia de los DBNs, son modelos totalmente no dirigidos y siempre con más de una capa oculta o variable latente, siendo cada una de ellas mutuamente independientes y condicionadas por las capas vecinas (las capas pares son condicionalmente independientes de las capas impares y viceversa). Las unidades que conforman estas capas suelen ser binarias pero también existe la posibilidad de que sean reales. El entrenamiento de estas redes es análogo al de las RBMs, por capas, aunque gracias a su estructura se consiguen mejoras en la tarea de clasificación, pudiendo ser todavía más acrecentadas si se realizan modificaciones a la hora de pre-entrenar y de entrenar. Como desventaja de esta variante destaca su complejidad temporal, algo poco deseable a la hora de trabajar con grandes conjuntos de datos.

3) Modelos de energía profunda (DEMs)

De las tres variantes aquí explicadas de las RBMs, los DEMs son las más recientes. A diferencia de las anteriores solo presenta una capa de unidades ocultas latentes, para así lograr un entrenamiento más rápido y eficiente, entrenando todas las capas a la vez y no capa a capa. Esto redunda en mejoras de clasificación cualitativas y cuantitativas.

2.2.3. Autocodificadores (AEs)

La forma más sencilla de entender un autocodificador es pensar en una red muy simple, formada por tres capas, una de entrada, una oculta y una de salida, y entrenada para copiar sus entradas a sus salidas. Lógicamente, su diseño no busca hacer una copia exacta. Por el contrario, la capa oculta impone restricciones para que se copie solo aquella información de entrada que sea relevante, priorizando unos aspectos de los datos, aquellos que resultan más útiles, sobre otros. Así, aunque a simple vista la salida parezca idéntica a la entrada, ésta ha sido reducida, se le han eliminado características que no eran relevantes. De forma más específica, para producir esta reconstrucción se sigue un proceso en dos etapas, una codificadora y otra decodificadora, y a medida que se ejecuta el proceso se va optimizando al minimizar el error de reconstrucción, para obtener finalmente la nueva función aprendida. Cabe resaltar que por el hecho de que el autocodificador busca copiar la entrada en la salida, ambas tienen que ser de las mismas dimensiones.

Esta breve descripción de la estructura de un autocodificador puede llevar al lector a recordar el funcionamiento de una red *feedforward*, y efectivamente pueden ser entrenadas de la misma manera. Además, a medida que se entrena al autocodificador hay que ir comparando las activaciones de la red original con las de la red reconstruida.

Tradicionalmente se usan para reducir la dimensión de las características o para el aprendizaje de las mismas, pero actualmente también se están empleando para tareas de recuperación de información y como modelos generativos de características. La generación de características es un paso de vital importancia cuando se busca detectar y clasificar lesiones en imágenes médicas, y es esencial realizarlo con criterio para obtener buenos resultados. Para ello, en los últimos años se han sustituido los autocodificadores simples por autocodificadores profundos (DAEs), pues como ya se ha explicado más capas pueden obtener más características, y tendrán un mayor potencial para averiguar las características discriminatorias y representativas de aquellos datos sin procesar. Para entrenar los DAEs primero se pre-entrenan con pesos iniciales que se aproximen a la solución final y luego se entrenan con una variante del algoritmo de *back-propagation.*

Cuando unas líneas más arriba se explicaba la estructura básica de un autocodificador, se hablaba de eliminar ciertas características de la entrada. Lograr este propósito implica que la dimensión de la capa oculta sea menor que la de las capas de entrada y salida, y es lo que se conoce como autocodificador incompleto (*sparse autoencoder*). Aunque esto es lo más común, también se puede tener un autocodificador sobrecompleto (*overcomplete autoencoder*), en el caso de que la capa oculta tenga una mayor dimensión que la capa de entrada. En esta situación y por mucho que se ejecute el algoritmo, no se obtendrá ninguna característica útil o saliente de los datos de entrada.

Para finalizar esta sección se describen las tres variantes más conocidas de este tipo de modelos:

1) Autocodificador disperso (*Sparse Autoencoders*, SAE)
Un autocodificador disperso es aquel que a la hora de entrenar asigna una penalización dispersa (*sparsity penalty*) a las capas ocultas, que se suma al error de reconstrucción. Su principal uso es aplicarlo a datos sin procesar para extraer y generar de características que luego serán empleadas en otra tarea, habitualmente en la de clasificación.

Las ventajas que proporciona esta arquitectura son tres. Primero hace a las categorías más fácilmente separables, por otro lado, logra que los datos complejos se interpreten de manera más fácil, y finalmente funciona de igual forma que el sistema de visión biológico, y por ello es de utilidad para aplicaciones en relación con el mismo.

2) Autocodificador con eliminación de ruido (*Denoising Autoencoders*, DAE)

De hacer de los autocodificadores modelos más robustos al ruido surgieron este tipo de arquitecturas. Su funcionamiento se basa en que, dada una copia de la entrada a la que se le ha añadido algún tipo de ruido, eliminan este ruido, logrando recuperar la entrada correcta de la versión dañada, en lugar de únicamente copiar la entrada. Esto se consigue al cambiar el término del error de reconstrucción en la matriz de coste.

3) Autocodificador contractivo (*Contractive Autoencoders,* CAE)

Los CAEs buscan, al igual que los DAEs, aprender representaciones más robustas, consiguiendo que el proceso de extracción de características sea resistente a pequeñas perturbaciones en la entrada. Esto se logra al añadir una penalización a la función del error de reconstrucción, con lo que se capturan mejor las direcciones de variación de los datos.

2.2.4. Codificación dispersa (*sparse-coding*)

Los modelos de codificación dispersa buscan, al igual que los autocodificadores, extraer características de unos datos de entrada para describirlos de forma completa. Son modelos lineales de aprendizaje no supervisados, que funcionan añadiendo ruido, normalmente gaussiano, a los datos de entrada, para obtener reconstrucciones de las mismas. El entrenamiento que se suele aplicar es por fases, alternando aquellas que codifican los datos con otras para la reconstrucción de los datos dada la codificación.

La principal ventaja de estos métodos es que no producen errores de generalización, y por lo tanto resultan ser mejores generalizadores cuando se usan como extractores de características, incluyendo los casos en los que se dispone de muy pocos datos etiquetados para entrenar. Esto es de gran utilidad al trabajar con imágenes médicas, dado que no siempre van acompañadas por etiquetas. Por el contrario, como desventajas se encuentran la gran cantidad de tiempo que emplea en hacer los cálculos y la dificultad de su etapa de entrenamiento.

La extracción de características es la principal aplicación de los modelos de codificación dispersa, ya que además de la ventaja de no requerir apenas de datos etiquetados, tiene otras muchas como por ejemplo: reconstruye mejor los descriptores al capturar las correlaciones entre descriptores similares; captura de manera eficaz las propiedades salientes de las imágenes; funciona del mismo modo que el sistema visual biológico; trabaja muy bien con los *patches* de imágenes, una estrategia de entrenamiento de redes profundas para detectar objetos de la que se hablará más adelante, por ser estos señales dispersas; y los patrones con características dispersas son más linealmente separables.

Para finalizar, y sin entrar en detalle, se mencionan algunos de los algoritmos más representativos de codificación dispersa.

1) SPM de codificación dispersa (ScSPM): Es una extensión de las ya mencionadas SPMs. Con ellas se consigue que el error de reconstrucción sea mucho menor, pero también se ignora la dependencia mutua entre las características locales, pues las trata por separado, lo cual no es ventajoso.
2) *Laplacian Sparse Coding (LSC):* Esta variante mejora a la ScSPM al seleccionar los centros de los cluster de forma que sean similares, logrando una solución de mayor robustez.
3) *Hyper-graph Laplacian Sparse Coding (HLSC):* El HLSC es una extensión del LSC donde la similitud entre las distintas instancias se define por un híper gráfico, consiguiendo con esta mejora mayor robustez.

2.2.5. Comparación entre modelos

Con el fin de comprender la clasificación que se ha realizado de las distintas técnicas de Aprendizaje Profundo, se resumen las propiedades de las cuatro categorías establecidas a modo de tabla comparativa, de forma general (no se tienen en cuenta hallazgos particulares).

Propiedades\Modelo	CNNs	RBM	Autocodificador	Codificación dispersa
Generalización	SI	SI	SI	SI
Aprendizaje no supervisado	NO	SI	SI	SI
Aprendizaje de características[1]	SI		SI	NO
Entrenamiento en tiempo real	NO	NO	SI	SI
Predicción en tiempo real[2]	SI	SI	SI	SI
Comprensión biológica	NO	NO	NO	SI
Justificación teórica[3]	SI	SI	SO	SI
Invarianza[4]	SI	NO	NO	SI
Conjunto de entrenamiento pequeño[5]	SI	SI	SI	SI

Tabla 1: Comparación entre los cuatro principales grupos de modelos de aprendizaje profundo.

3. Aplicaciones del Aprendizaje Profundo

En este apartado se detallan las aplicaciones y logros de los algoritmos de Aprendizaje Profundo en visión por ordenador, y a continuación se explican con detalle las mismas tareas pero enfocadas al análisis de imágenes médicas (MIA), donde se apreciará el claro predominio de uso de las CNNs.

La visión por ordenador ha sido, desde los orígenes del Aprendizaje Profundo una de las tareas con más investigación, por ser muy sencilla para los humanos pero muy costoso para los ordenadores. En concreto las tareas donde más se ha trabajado han sido el reconocimiento de objetos y el reconocimiento de caracteres ópticos. La investigación en este campo se ha visto motivada a su vez por sus múltiples aplicaciones, que van desde el reconocimiento facial hasta la creación de nuevas habilidades visuales [2].

La mayoría de algoritmos empleados se centran en el reconocimiento de objetos o en su detección, lo que significa informar qué objetos están presentes en una imagen, anotar los objetos identificados en la imagen de alguna manera (i.e.

[1] Aprendizaje de características → capacidad de aprender automáticamente características basadas en un conjunto de datos.
[2] Entrenamiento en tiempo real, predicción en tiempo real → se refieren a la eficiencia de los procesos de aprendizaje e inferencia, respectivamente.
[3] Comprensión biológica, justificación teórica → Hacen referencia a si el enfoque tiene bases biológicas significativas o fundamentos teóricos, respectivamente.
[4] Invarianza → Si el enfoque ha sido robusto a transformaciones tales como rotación, escala y traducción.

cajas delimitadoras, contornos, etc.), transcribir una secuencia de símbolos desde una imagen, o etiquetar cada píxel en la imagen con la identidad del objeto al cual pertenece. Son distintas aproximaciones, pero todas ellas con el fin de detectar una forma u objeto.

El punto común del que parten todas las aplicaciones de visión por ordenador es buscar procesar de una serie de imágenes. Para procesar una imagen con un fin, primero se debe adecuar y/o pre-procesar, buscando así lograr resultados siempre mejores. En este trabajo se hablará exclusivamente de los tipos de pre-procesamiento comunes en imagen médica, y esto se hará en el Capítulo 4.

3.1. Aplicaciones del Aprendizaje Profundo en visión por ordenador

A continuación se resumen las principales aplicaciones de los métodos de Aprendizaje Profundo en visión por ordenador, comenzando por la tarea más popular, la de clasificación. No se detallarán en profundidad puesto que esto se hará para su aplicación en particular al análisis de imagen médica, líneas más adelante.

3.1.1 Clasificación de imágenes

La clasificación de imágenes consiste en dado un conjunto de imágenes, etiquetar a cada una de ellas con una probabilidad de que pertenezca a una clase o a otra.

Las CNNs se declararon en el 2014 como las arquitecturas más eficientes para esta tarea, cuando la mayoría de participantes de ILSVRC 2014 [24]las escogieron como base para sus modelos de clasificación, obteniendo buenos resultados. Con el modelo de SPP-*net* se consiguió eliminar la restricción de que la imagen de entrada tuviera un tamaño fijo, lo que evidentemente fue un gran avance y mejoró la precisión de todo tipo de arquitecturas basadas en CNNs. Otra característica que hizo a las redes mejores fue el aumentar su profundidad, como se ha visto con *GoogLeNet*, aunque estos últimos modelos son más susceptibles al *overfitting* y al *underfitting* en el caso de tener pocos datos de entrenamiento o poco tiempo, problemas cuya solución todavía se está buscando, aunque todo apunta a usas la técnica del *Deep-Image*, que permite por un lado usar imágenes de distintos tamaños y por otro aumentar el número de datos.

3.1.2 Detección de objetos

Es una tarea que va unida la de clasificación de objetos, y para esta también se usa una imagen como entrada y se suelen estimar las etiquetas de las clases de objetos que tiene esa imagen; pero además se busca obtener la posición de los objetos. Por lo tanto, no solo da información acerca de la existencia de una clase, si no también acerca de su localización. Para lograr el localizar los objetos se emplean distintas estrategias, siendo la más popular la de usar una ventana de detección que se solape con el objeto al menos en un 50%.

En tareas de visión por ordenador no médicas, con imágenes naturales, el conjunto de datos *PASCAL VOC*, con 20 clases, es muy popular y suele emplearse para la evaluación de la tarea de detección.

Por compartir muchas características en el proceso, y a la vista de los buenos resultados que obtenían, se optó por usar CNNs para la detección de objetos. De aquí surgió la arquitectura *DetectorNet* [25], análoga a *AlexNet* pero con una capa

de regresión como última capa. Tras ella apareció *DeepMultiBox,* para el manejo de múltiples instancias del mismo objeto en una misma imagen.

Como se avanzaba unas líneas más arriba, el esquema general para lograr una detección de objetos exitosa es generar un grupo con múltiples cajas candidatas y clasificarlas usando una CNN, en concreto una RCNN, que se recuerda consistía en una CNN seguida de un SVM lineal, para que así las propuestas de candidatos sean selectivas. Las RCNNs son la base de la mayoría de algoritmos empleados en esta tarea de detección, y los constantes estudios que buscan mejorar su desempeño suelen hacerlo centrándose en acelerar los procesos de entrenamiento y prueba o en mejorar la precisión de la red. Los primeros pretenden obtener detecciones de objetos de forma más rápida, pues al generar muchos candidatos son computacionalmente costosos al tener que procesar cada uno de ellos por separado. Los segundos, de mayor interés para este estudio, buscan mejorar la precisión de la localización de los objetos, y proponen emplear otras técnicas para lograrlo. Esto es algo de vital importancia en la detección de lesiones y tumores para su posterior clasificación o segmentación.

El gran reto a la hora de detectar objetos es la dificultad de obtener imágenes etiquetadas para un gran número de categorías, ya que no es barato ni fácil lograr que las etiquetas de las imágenes estén a nivel de regiones o de píxeles. Este problema se acrecienta todavía más en bases de datos de imágenes médicas. Las soluciones que se están explorando proponen usar arquitecturas más profundas, en particular el algoritmo de Adaptación de Detección Profunda (DDA) y los modelos *ConceptLeaner* [26] y *BabyLearning* [27]. Estas dos aproximaciones no necesitan hacer una anotación masiva de conceptos visuales para la detección de los objetos, lo cual es una gran ventaja pues conseguir imágenes no explícitamente anotadas por humanos pero que compartan algunas características, que es lo que emplea *ConceptLearner*, es algo poco costoso; y *BabyLearning* solo precisa unas pocas muestras etiquetadas, entre muchas sin etiquetar, para cada categoría de objetos.

3.1.3 Reconocimiento de imágenes

El reconocimiento de imágenes consiste en buscar aquellas imágenes que contengan un objeto o una escena similar a la imagen de entrada.

La estrategia más empleada son modelos basados en CNNs, motivado una vez más por los buenos resultados de *AlexNet* en clasificación de imágenes. Estos buenos resultados sugieren que las características que resultan en las primeras capas de una CNN para clasificación de imágenes pueden ser buenos descriptores para la clasificación de las mismas, y los resultados obtenidos lo demuestran. Sin embargo, el decidir qué capa de la red es mejor utilizar, cuál tiene un mayor impacto, es algo que permanece abierto, además de ser cambiante según el conjunto de datos que se esté empleando.

Lo más típico es encontrar CNNs con variaciones como CNNs profundas pre-entrenadas con grandes conjuntos de datos, utilizadas para la extracción de características en tareas CBIR (identificación de imágenes basada en contenido) y luego re-entrenarlas con aprendizaje por similitud. También se suele encontrar otra variante que extrae primero trozos de la imagen que se parezcan al objeto con un detector de objetos genérico y luego extrae características de cada trozo del objeto

con el modelo preentrenado de *AlexNet*. Estrategias de este tipo se usan para conjuntos de imágenes médicas, y está demostrado que la precisión aumenta.

Además de CNNs para el reconocimiento de imágenes también se usan descriptores holísticos, en los cuales la imagen completa es mapeada a un solo vector con un modelo de CNN.

3.1.4 Segmentación semántica

La segmentación semántica asigna una etiqueta o categoría a cada píxel de una imagen. Para ello se emplean, de nuevo, CNNs, pues son capaces de hacer estas predicciones a nivel de píxel en conjuntos de datos muy grandes. Algo fundamental en la segmentación semántica es tener una máscara de salida con una distribución espacial 2D. Los principales métodos basados en CNNs que se emplean son los tres que siguen.

- **Segmentación basada en la detección.** Consiste en segmentar las imágenes a partir de las ventanas candidatas que resultan de la detección de objetos. Para el primer paso se usan RCNN y SDS y para el segundo se emplean aproximaciones tradicionales de aprendizaje de máquina. Su principal desventaja es el gran coste de detectar del objeto. Otro método para evitar extraer regiones de imágenes sin procesar es el CFM (*Convolutional Feature Masking*), que extrae las propuestas directamente de los mapas de características, lo cual resulta muy eficiente. En ambos casos hay que tener cuidado pues los errores causados por las distintas propuestas y por la detección del objeto suelen propagarse en el paso de segmentación.

- **Segmentación basada en FCN-CRFs.** Es una estrategia muy popular para la segmentación semántica, existiendo distintas variantes con pequeñas modificaciones que obtienen buenos resultados.

- **Anotaciones débilmente supervisadas.** El concepto de anotaciones débiles se refiere a que la información que se tiene de los datos no es a nivel de píxel, si no a nivel de imagen, por ejemplo. Estos métodos tienen buenos resultados si se combinan un pequeño número de imágenes del primer tipo con un gran número de imágenes del segundo.

3.1.5 Estimación de la pose humana

La estimación de la pose humana tiene como objetivo localizar las articulaciones humanas a partir de imágenes inmóviles o a partir de secuencias de imágenes. Esto es importante para tareas como la videovigilancia, el análisis del comportamiento humano o la interacción hombre-máquina (HCI), etc. Resulta una tarea muy desafiante debido a la amplia variabilidad en las apariencias humanas, a los fondos complicados de las imágenes, y a factores de ruido como la iluminación o la escala.

A pesar del interés que puede suscitar esta tarea, se queda fuera de las aplicaciones médicas de estos algoritmos. Por ello solo se mencionarán un par de técnicas que se emplean para estimar la pose humana en imágenes estáticas, que son las de procesado holístico, como el *DeepPose*, que procesa la imagen de forma global, y las de procesado basado en partes, que proporcionan mejores resultados al detectar las distintas partes del cuerpo de manera individual y luego incorporar la información espacial de las mismas.

Finalmente, mencionar que la línea de trabajo a seguir es lograr incorporar características de movimiento para que así puedan usarse estas técnicas también en vídeos, y no solo en imágenes estáticas.

Antes de pasar a tratar con detalle las aplicaciones médicas del Aprendizaje Profundo, se resumen brevemente otras cuatro tareas genéricas del campo de visión por ordenador. Estas son:

- Reconocimiento de voz. Consiste en mapear una señal acústica con una declaración de un hablante en su correspondiente secuencia de palabras. Es una tarea con historia, que empezó alrededor de 1980, pero no fue hasta el 2009 cuando se comenzó a utilizar modelos de Aprendizaje Profundo no supervisado, en concreto RBMs, que luego quedaron en desuso siendo reemplazadas por CNNs y RNNs profundas.

- Procesado del Lenguaje Natural (NLP): Consiste en el uso del lenguaje humano, en diferentes idiomas, por parte de un ordenador (i.e. en máquinas traductoras). En general para desempeñar este tipo de tareas se pueden usar redes neuronales genéricas, pero para lograr resultados realmente buenos conviene aplicar ciertas estrategias, como por ejemplo de procesamiento de datos por secuencias (de palabras, de caracteres o de bytes).

- Sistemas de recomendación: Todavía en fase de investigación y desarrollo, los sistemas de recomendación son sistemas que hacen recomendaciones a usuarios potenciales o consumidores, para lo cual se basan en la asociación entre un usuario y un producto y prediciendo la probabilidad de que se compre ese producto. Este problema se modela como un problema de aprendizaje supervisado, pues se tiene información sobre el usuario y sobre el producto, y a partir de esa información el algoritmo tiene que hacer una predicción, por regresión o clasificación probabilística.

- Representación de conocimiento, razonamiento y respuestas a preguntas: Es otra aplicación que todavía está en fase de investigación y desarrollo, pues todavía no existen sistemas que capten adecuadamente las relaciones entre palabras y hechos.

3.2. Aplicaciones en imagen médica

Tras haber revisado brevemente las aplicaciones en visión por ordenador, el lector debe de estar ya concienciado del potencial que tienen las técnicas de Aprendizaje Profundo. Llegados a este punto, se quiere hablar de cómo se pueden aplicar, y de cómo se están aplicando, estos algoritmos para analizar imágenes médicas, concienciándole así de las múltiples ventajas que proporcionan. Para ello se hará una revisión del estado del arte tanto de las tareas (Figura 5) como de los órganos en los que se están aplicando las CNNs y demás algoritmos. Tras ello se prestará total y única atención a sus aplicaciones relativas al cáncer de mama.

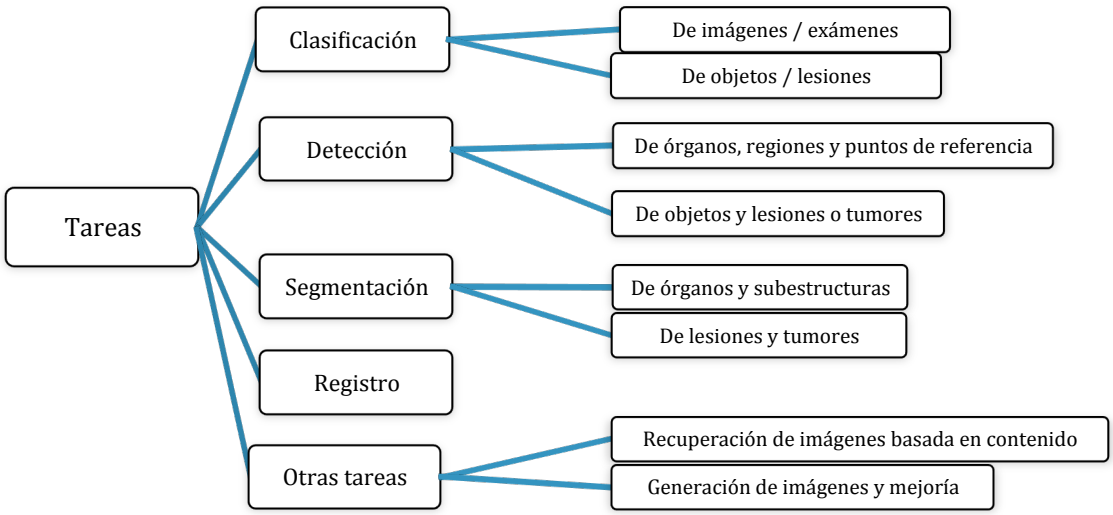


Figura 5: Esquema de las tareas de análisis de imagen médica donde se aplican métodos de Aprendizaje Profundo.

a)Clasificación:

"Acción de clasificar - Ordenar o dividir un conjunto de elementos en clases a partir de un criterio determinado - "

- Clasificación de imágenes (exámenes):

La clasificación de imágenes es una de las tareas médicas a las que más ha contribuido el Aprendizaje Profundo hasta el momento, por no decir a la que más. Por ello, se encuentran multitud de trabajos en imágenes de distintas modalidades (CT, MRI, US) y para tipo todo de partes del cuerpo (i.e. cerebro, pulmones, mamas, retina, etc.) [4].

El esquema habitual de la clasificación de imágenes comienza por tener como entrada un conjunto de muchas imágenes, lo que sería un examen realizado a un paciente, y como salida una única variable de diagnóstico, por ejemplo, decir si una cierta enfermedad está presente o no. Cabe resaltar el hecho de que el conjunto de imágenes (el examen) se toma como una sola muestra, lo que hace que el conjunto de datos se reduzca enormemente en comparación con los formados por imágenes naturales no médicas.

A continuación, a este examen de entrada se le aplican distintas técnicas que suelen basarse en Aprendizaje de Transferencia, por medio de redes pre-entrenadas en conjuntos de imágenes naturales para la extracción de las características, o bien ajustando estas mismas redes pre-entrenadas a los propios datos médicos que se tienen. En cualquier caso, resulta ventajoso en el sentido de que evita el paso de entrenamiento de la red.

Las arquitecturas que se emplean típicamente para desempeñar esta tarea son SAEs, RBMs, ambas entrenadas de manera no supervisada, y CNNs, siendo este último tipo de red la más utilizada en los últimos años (un 76,6% de los artículos publicados entre 2015 y 2017 lo confirman [4]). Además, el uso de CNNs ha demostrado tener un desempeño muy bueno, llegando a desafiar la precisión de expertos humanos.

Recientemente también se pueden encontrar, como estrategia alternativa al uso de redes pre-entrenadas, artículos donde se emplean redes propias, construidas por los propios autores de los artículos. Esta estrategia de diseño propio es precisamente lo que se busca hacer en este trabajo, como se plasma en el Capítulo 4.

- Clasificación de objetos (lesiones):

Otro tipo de clasificación es aquella realizada sobre una única imagen, y no a un conjunto, con el objetivo el clasificar pequeñas partes de dicha imagen en dos o más clases. Para ello, las partes a clasificar ya tienen que estar identificadas, bien por una etapa previa de la red o manualmente. Como partes de la imagen u objetos se hace referencia a todo tipo de estructuras anatómicas, lesiones, tumores, etc.

Para que esta tarea sea realizada con éxito es de gran importancia tener tanto información local del propio objeto a clasificar como información del contexto global, de lo que rodea a dicha lesión. Se puede entender esta necesidad haciendo la comparación con un médico, que si solo ve un trozo de una imagen no va a ser capaz de realizar un diagnóstico preciso. Sin embargo, tener ambos tipos de información es algo difícil a la hora de trabajar con arquitecturas de Aprendizaje Profundo, y por ello se propone como solución combinar arquitecturas, normalmente una CNN con otra CNN o una CNN con una RNN, de forma que se puedan procesar grandes cantidades de información, grandes imágenes.

Otro hándicap a la hora de clasificar lesiones es el incorporar la información 3D de las imágenes. Muchas de las redes que hay hasta ahora han sido desarrolladas específicamente para problemas de visión por ordenador, y por lo tanto para imágenes 2D, y no pueden manejar de forma directa información 3D. El integrar información 3D es algo interesante porque está demostrado que mejora en gran medida la tarea de clasificación, y se puede hacer empleando RBMs, SAEs, y CSAEs pre-entrenados de forma no supervisada con autocodificadores dispersos, y también con CNNs entrenadas de extremo a extremo (*End-to-end*).

Dentro de esta tarea tiene lugar también mencionar el *Multiple Instance Learning* (MIL) que, si se combina con técnicas de Aprendizaje Profundo, mejora los resultados de la clasificación, sobre todo en aquellos casos donde es muy costoso generar datos para el entrenamiento porque se requiere hacer anotaciones en los objetos. Aun no estando muy extendida, esta combinación de métodos se espera que sea muy exitosa en los próximos años en imagen médica, evitando la necesidad de disponer de información médica anotada.

b) Detección

"Acción de detectar – Captar o notar la presencia de una persona, una cosa o un fenómeno - "

- Localización de órganos, regiones y puntos de referencia (o *landmarks*)

A la hora de hablar de clasificación de objetos se ha mencionado que estos objetos ya tienen que estar previamente identificados y localizados. Así, se puede ver que la detección de una determinada estructura anatómica es un paso muy importante en el análisis de imagen médica, y si no es realizado de una forma correcta, causará problemas como una segmentación incorrecta de la estructura de interés o como dificultar el flujo clínico de una terapia o de una intervención.

El localizar estructuras en imágenes médicas suele implicar analizar los volúmenes en 3D, para lo cual se siguen tres líneas diferentes: interpretar el volumen en 3D como una serie de planos 2D ortogonales compuestos entre ellos; identificar la ROI, que será la región anatómica de interés, por medio de CNNs preentrenadas y RBMs, y por clasificación; y modificar el proceso de aprendizaje para que la red prediga directamente la localización de la estructura de interés. De estos tres procesos el segundo, el localizar las estructuras de interés en 2D tratando la tarea como si fuera un proceso de clasificación, es el más empleado, mientras que el último de ellos es el más complicado, pero del cual en un futuro se espera llegar a tener mejores resultados de localización.

Para concluir, comentar también la aplicación de métodos de Aprendizaje Profundo para trabajar con vídeos médicos, lo cual muestra el alto potencial de las RNNs para localizar estructuras en un dominio temporal.

- Detección de objetos (lesiones)

La detección de las regiones de interés, de las lesiones, es uno de los puntos clave en cualquier diagnóstico médico, además de ser una de las tareas que más trabajo da a los especialistas. Lo más habitual es que en una imagen haya más de una pequeña lesión, y consecuentemente en la tarea de detección tienen que localizarse e identificarse todas y cada una de ellas.

Para esta tarea se tienen los conocidos como Sistemas de Ayuda a la Detección o CADe, en los cuales se trabaja constantemente para mejorar su precisión en la detección, disminuir el tiempo de lectura de las imágenes, y en definitiva asesorar a los expertos y ayudarles en su labor diaria.

La forma típica en la que funcionan estos sistemas es la siguiente: primero se hace una clasificación de todos los píxeles o vóxeles de la imagen empleando una CNN, y a continuación se aplica algún tipo de pre-procesado para obtener todos los objetos candidatos. La arquitectura de estas CNNs y la metodología que siguen es análoga a la empleada en la clasificación de objetos, pues son tareas equiparables. Del mismo modo que en la clasificación, en esta otra también es útil incorporar información de contexto, para lo cual se están empleando CNNs *multi-stream*.

Aún teniendo muchos aspectos en común, obviamente existen otros muchos en los que la detección de objetos difiere de la clasificación de los mismos, como por ejemplo el hecho de, en la tarea de detección, al ser todos los píxeles clasificados o bien como candidatos o como no-candidatos, siempre va a haber muchos más píxeles correspondientes a la clase no-candidato, que además suelen ser píxeles muy sencillos de discriminar (comparten características más similares). Si se suma esta relativa facilidad de clasificación con la gran diferencia en proporción de unos sobre otros, resulta que el algoritmo termina centrándose más en clasificar aquellos píxeles que no son de interés que aquellos que sí lo son, pues los píxeles de la lesión en particular terminan siendo un reto. Por ello es muy importante en esta tarea aplicar determinadas técnicas para lograr un balanceado entre ambas clases, ayudando así que el algoritmo se centre en ambas por igual.

Para concluir esta sección se podría decir que los aspectos en los que la detección de objetos o lesiones difiere de la detección son precisamente aquellos que suponen un desafío actualmente.

c) Segmentación

"Acción de segmentar – Cortar o partir algo en segmentos -"

- Segmentación de órganos y subestructuras

De forma técnica, la segmentación se suele definir como el identificar el grupo de vóxeles que constituyen el contorno o el interior del objeto de interés.

En imágenes médicas, la segmentación de órganos y otras subestructuras es esencial para poder hacer un análisis cuantitativo de parámetros clínicos en relación con la forma y con el volumen de las mismas estructuras (i.e. para análisis cerebrales o cardíacos). Además, la segmentación es un primer paso en los sistemas de detección por ordenador, cobrando todavía una mayor importancia.

Por esta importancia de realizar bien la segmentación de las estructuras de interés existen multitud de modelos que proponen distintos enfoques para abordar el problema, desde arquitecturas CNNs específicas hasta RNNs.

De las CNNs específicas cabe mencionar *U-Net*, publicada en 2015 [28]. Sus principales novedades son la combinación, en igual cantidad, de capas de *upsampling* y de capas de *downsampling,* y la presencia de conexiones entre capas opuestas de convolución y de deconvolución, que permiten concatenar características de estas capas. Con estas dos mejoras, y desde una perspectiva de entrenamiento, esto hace que las imágenes puedan ser procesadas por la red en un único paso hacia delante, resultando en el mapa de segmentación directamente. Además, gracias a esta estructura, la *U-net* tiene en cuenta el contexto global de la imagen, algo que como se ha comentado de forma continuada es una gran ventaja en comparación con las CNNs estándar. A partir del modelo de esta *U-Net* otros autores han implementado otras arquitecturas que proporcionan ciertas mejoras.

Por otro lado, para evitar, o al menos disminuir, la computación redundante provocada por el uso de ventanas deslizantes para ir clasificando los píxeles, hay autores que proponen usar fCNNs. El desempeño de estas redes es muy satisfactorio pues pueden ser aplicadas a múltiples objetivos a la vez (i.e. con la misma fCNN entrenada se segmentan de distintas imágenes tanto el cerebro en MRIs, el músculo pectoral en MRIs y las arterias coronarias en CTAs) [29].

Finalmente, el principal desafío al que se enfrentan a día de hoy las técnicas de segmentación es reducir la tasa de vóxeles incorrectamente clasificados, para lo cual recientemente se ha propuesto combinar las fCNNs con Modelos Gráficos (MRFs) y con Campos Aleatorios Condicionales (CRFs), que consiguen refinar la salida de la tarea de clasificación.

- Segmentación de lesiones

La segmentación de lesiones combina tanto los desafíos de la detección de objetos como los de la segmentación de órganos y subestructuras.

Con la detección de objetos comparten las características del desequilibrio entre clases, y que es necesario tener información tanto local como del contexto global para que la localización de la lesión sea precisa, y para esto último se usan redes del tipo *U-net* y sus derivadas.

Por ello, la segmentación de lesiones combina los enfoques de la detección de objetos con los de la segmentación de órganos, y cualquier avance realizado en estos dos campos se propagará con probabilidad a la segmentación de lesiones.

d) Registro

"Acción de registrar – Dejar registro impreso de imágenes, sonidos en un disco, en una cinta magnética o en otro soporte material para poderlos reproducir. -"

El registro de imágenes médicas, también conocido como la alineación espacial de imágenes, consiste en dada una imagen inicial, aplicarle una transformación de coordenadas, obteniendo una imagen final. A menudo se asume un tipo específico de transformación (i.e. no paramétrica) y se emplea una métrica predeterminada (i.e. la norma L2).

Aun no siendo tan empleadas las redes profundas para esta tarea, pueden aportar muchos beneficios a la hora de obtener el mejor registro posible. Para ello se usan dos estrategias; utilizar las redes para estimar una medida de similitud entre dos imágenes dadas, y a partir de ésta aplicar una estrategia de optimización iterativa; o predecir directamente los parámetros de transformación utilizando redes de regresión profunda.

Todavía no existen muchos artículos de investigación sobre el tema y los pocos existentes tienen un enfoque claramente distinto, por lo que afirmar que un método es más prometedor que otro no sería lo más adecuado por ahora.

e) Otros usos en imagen médica

- Recuperación de imágenes basadas en su contenido (*Content Based Image Retrieval,* CBIR*)*

La CBIR consiste en descubrir conocimiento en bases de datos masivas, lo cual en análisis médico es de utilidad a la hora de identificar casos similares, entender trastornos raros y, en última instancia, mejorar la atención al paciente.

El principal reto en el desarrollo de los métodos CBIR es extraer representaciones de características efectivas de la información a nivel de píxel y asociarlas con conceptos significativos, tarea para la cual los modelos de CNN profundos funcionan de forma efectiva por su capacidad para aprender características complicadas a múltiples niveles de abstracción. Así, todos los trabajos hasta la fecha utilizan CNNs pre-entrenadas.

Por ahora, los métodos de aprendizaje profundo no han tenido muchas aplicaciones exitosas en esta tarea, aunque se espera que en un futuro cercano esta situación cambie.

- Generación y mejora de imágenes

La generación y mejora de imágenes abarca tareas más bien de pre-procesado de las mismas, como la eliminación de elementos obstructivos en las imágenes, la normalización de las imágenes, la mejora de la calidad de la imagen, el completar datos y el descubrimiento de patrones.

Para la generación de imágenes, se usan CNNs 2D o 3D para convertir una imagen de entrada en otra. Estas arquitecturas suelen carecer de las capas de agrupación y por ello se requiere realizar un entrenamiento con un conjunto de

datos en el que se incluyan la entrada y la salida deseadas, y en el que se definan las diferencias entre ellas, así como la función de pérdida. Un ejemplo de su uso en imagen médica se plasma en [30] donde se generaron una serie de imágenes y se emplearon en un CAD para el diagnóstico de Alzheimer en casos donde no se tenían los datos originales, debido a que habían sido adquiridos o a que no se encontraban disponibles.

Esta aplicación es solo un ejemplo de que las CNN son muy útiles para inferir información que falta. Sin embargo, en las otras tareas de pre-procesado mencionadas (normalización, degradación) todavía no se ha encontrado que su uso aporte mejores significativas.

- Combinación de datos: Imágenes con informes de texto

El tener una imagen médica asociada a un informe es algo de gran utilidad pues ayuda a los especialistas en sus tareas. El poder realizar esto de forma automática sería algo muy provechoso. Por ahora, las dos tareas que se realizan son el usar estos informes y aprovecharlos para mejorar la precisión de la clasificación de las imágenes, y el generar informes de texto a partir de las imágenes.

Así pues, tras haber revisado el estado del arte de las aplicaciones de las Redes Neuronales en el campo de la imagen médica, resulta de mayor interés ver en qué áreas del cuerpo humano se están aplicando con un mayor éxito. De esta manera, se agrupan los usos de Aprendizaje Profundo por órganos en las secciones a continuación.

- Cerebro

Para analizar imágenes cerebrales la técnica más empleada son las DNN, con distintas aplicaciones [4]. Entre todas ellas destaca la clasificación de la enfermedad de Alzheimer, por el gran número de estudios que tiene asociados, y la segmentación del tejido cerebral y de las estructuras anatómicas (i.e. del hipocampo). También se trabaja mucho en la detección y segmentación de lesiones cerebrales como los tumores, las lesiones de materia blanca, las lagunas, y los micro-sangrados.

La mayoría de estos métodos funcionan aprendiendo primero a mapear *patches* locales a representaciones, y luego estas representaciones a etiquetas de clasificación.

Por otro lado, cabe mencionar que a pesar de que las imágenes cerebrales son volúmenes en 3D, la mayoría de los métodos funcionan en 2D, y consecuentemente analizando los volúmenes 3D rodaja a rodaja. Esto se hace así para que el coste computacional sea menor, o porque en muchos casos los volúmenes resultan demasiado gruesos para su análisis. Aun así, las publicaciones más recientes han empleado redes que funcionan en 3D [31] [32].

En cuanto a la modalidad de imagen empleada, en casi todos los casos se usan imágenes de RM cerebral, pero se espera que otras como el CT y los US comiencen a ser empleados.

- Ojos:

Los algoritmos de Aprendizaje Profundo se aplican a la comprensión de la imagen oftálmica desde hace muy poco. Lo más destacable y trabajado es el empleo de CNNs simples para analizar la retinografía de color (*color fundus imaging*, CFI). De este análisis surgen distintas aplicaciones; la segmentación de estructuras anatómicas, la segmentación y detección de anomalías de la retina, el diagnóstico de enfermedades oculares y la evaluación de la calidad de la imagen ocular.

Una aplicación que ha alcanzado particular éxito es la detección de retinopatía diabética con CNNs, alcanzando estas redes mejores resultados que cualquier otro método e incluso mejores que los obtenidos por expertos humanos [33].

- Torso

De todas las aplicaciones que pueden resultar de analizar imágenes torácicas, tanto de RX como de CT; la de detección, caracterización y clasificación de nódulos es en la que más se trabaja. Para abordarla se emplean estrategias como añadir características obtenidas por redes profundas a conjuntos de características ya existentes, y se mide su desempeño comparando su precisión con la obtenida por enfoques clásicos de ML, que emplean solo los conjuntos de características ya existentes.

En CT lo más habitual es la detección de patrones de texturas indicativos de enfermedades pulmonares intersticiales; mientras que en las radiografías predomina la tarea de detección de múltiples enfermedades en el torso empleando un solo sistema, algo conseguido por más de un grupo de trabajo. De estas imágenes radiográficas, el examen más común es la radiografía de tórax. Varias obras utilizan un gran conjunto de imágenes con informes de texto para entrenar sistemas que combinan CNNs con RNNs para el análisis de imagen y de texto, respectivamente [4].

- Microscopía y patología digital

La creciente disponibilidad de imágenes de grandes dimensiones (WSI) de muestras de tejidos ha hecho que la patología digital y la microscopía se conviertan en un área de gran interés para aplicar técnicas de Aprendizaje Profundo. Actualmente se usa para tres aplicaciones principalmente; para la detección, segmentación o clasificación de núcleos, para segmentación de órganos grandes, y para detección y clasificación de la enfermedad de interés en la lesión o en la WSI.

También se usan técnicas de aprendizaje profundo para normalizar de imágenes histopatológicas, destacando la normalización de color de las mismas.

El desarrollo de técnicas de patología digital computarizada se ha fomentado por los muchos desafíos que han surgido en el área de la patología digital, como la segmentación 2D de procesos neuronales, la detección de la mitosis, la segmentación de distintas glándulas, y el procesamiento de muestras de tejido de cáncer de mama. Para abarcar todos estos retos se ha decidido emplear algoritmos basados en CNNs.

En relación con el siguiente órgano a estudiar, las mamas, cabe destacar que en 2016 en el Tumor Proliferation Assessment Challenge (TUPAC) se propuso

detectar la mitosis en el tejido canceroso de mama y predecir la clasificación del tumor con WSIs. El sistema de mayor rendimiento, en todas las tareas [34], funcionaba en tres pasos: encontrar las regiones de alta densidad celular, usar una CNN para detectar la mitosis en las ROIs, y finalmente convertir los resultados del paso anterior en un vector de características para cada WSI, para luego emplear un SVM que calculaba puntuaciones de proliferación tumoral y de datos moleculares.

- Imágenes cardíacas

El Aprendizaje Profundo se ha aplicado a muchos aspectos del análisis de imágenes cardiacas. La modalidad más usada para ello es la RM, y la tarea más se estudiada es la segmentación del ventrículo izquierdo; si bien existen muchas otras aplicaciones, como la segmentación de otras estructuras, el seguimiento de lesiones, la clasificación de imágenes, la evaluación de la calidad de la imagen, la puntuación automatizada para los niveles de calcio, y el seguimiento de la línea central en la arteria coronaria, entre otras.

La mayoría de los trabajos se basan en CNNs 2D simples, y para ello analizan los datos 3D, y a menudo los 4D, rodaja a rodaja. Existe alguna excepción donde se emplean CNNs para volúmenes 3D. También se pueden encontrar artículos que usan DBN, pero solo para la etapa de extracción de características. Finalmente resultan de interés dos artículos diferentes [35] [36] que combinan, en ambos casos, CNNs con RNNs, obteniendo resultados interesantes.

Un punto a favor de los trabajos realizados sobre imágenes cardíacas es que la mayoría utilizan BBDD que están públicamente disponibles, algo que por desgracia no sucede para las imágenes de muchas otras estructuras.

- Abdomen

Los estudios relativos al abdomen buscan localizar y segmentar los órganos que se encuentran bajo él, principalmente el hígado, los riñones, la vejiga y el páncreas. De forma más concreta otros trabajos abordan la tarea de segmentación de tumores, en particular hepáticos. La modalidad de imagen principalmente empleada es la TC para todos los órganos, a excepción de la RM en el caso de los análisis de próstata.

Si se tuviera que destacar un área por el amplio número de investigaciones que parten de ella ésta sería el colon [4], pues es el único órgano bajo el abdomen donde se ha encontrado más de una aplicación del Aprendizaje Profundo, aunque siempre de la misma manera; usando una CNN como extractor de características y utilizando estas características para la clasificación.

- Músculo-esquelético:

Las imágenes músculo-esqueléticas también se han analizado en varias ocasiones empleando algoritmos de aprendizaje profundo, en concreto para la segmentación e identificación del hueso, de articulaciones y de anomalías asociadas a tejidos blandos, todo ello en muy distintas modalidades de imagen.

- Mamas

Aunque la detección del cáncer de mama en imágenes tanto mamográficas como de otras modalidades es una de las tareas donde se pueden encontrar más trabajos de modelos de Aprendizaje Profundo, no se va a tratar este tema aquí, sino que se hará

más adelante, el Capítulo 4, pues se quiere abordar el problema de analizar las imágenes de las mamas con todo el detalle, por ser uno de los objetivos de este trabajo.

- Otras:
Finalmente, se agrupan aquellos trabajos dirigidos hacia otras aplicaciones médicas, cuyo desarrollo por ahora está algo más retrasado que los anteriores. De ellas destacan fundamentalmente las aplicaciones obstétricas, donde los trabajos buscan lograr una selección automática de la imagen apropiada de toda la secuencia que proporcionan los US, y las dermatológicas, que emplean imágenes dermoscópicas. De éstas últimas y hasta hace poco, los esfuerzos se han centrado en el diagnóstico del cáncer de piel a partir de fotografías por su alto grado de dificultad, pero últimamente muchos estudios han optado por trabajar solo con imágenes obtenidas con cámaras especializadas, consiguiendo a partir de enfoques con redes neuronales profundas resultados más que prometedores.

A modo de conclusión, cabe mencionar el hecho de que cada vez se está trabajando en arquitecturas de Aprendizaje Profundo que puedan aplicarse sin modificaciones a distintas tareas, haciendo así de las redes neuronales modelos versátiles y fácilmente generalizables. Esto se ha conseguido ya en algunos trabajos, que obtienen resultados competitivos pre-entrenando arquitecturas con imágenes de un dominio completamente diferente al médico, para el que se quieren aplicar las redes.

4. Diseño de una metodología para el análisis de mamografías

Como se ha comentado anteriormente, el Aprendizaje Profundo existe desde hace mucho más de lo que se piensa, si bien no se vio explotado hasta estos últimos años debido al hecho de que no se disponían de bases de datos públicas lo suficientemente amplias como para trabajar con él, y a que los ordenadores que había hasta el momento no eran lo suficientemente potentes, si bien ambos hándicaps han sido superados con las mejoras computacionales y la buena labor de recolección de datos para generar grandes bases de datos. Concretamente, el primer artículo relacionado con aplicar el Aprendizaje Profundo a la medicina fue publicado en 1996 [37] y tenía como objeto clasificar el tejido del pecho como cancerígeno o no en función de su textura. En los últimos años, el interés por el cáncer de mama ha sido creciente, por el gran número de afectados por un lado, y por otro por la gran cantidad de imágenes que se tienen, lo que ha hecho que surjan multitud de estudios que proponen técnicas y algoritmos diferentes para su detección y clasificación.

Pero, ¿qué es exactamente el cáncer de mama?

4.1. El cáncer de mama

El cáncer de mama es hoy en día es el segundo tipo de cáncer más frecuente, con una incidencia del 12%. En el año 2012 se diagnosticaron cerca de 1,7 millones de nuevos casos en el mundo, y solo en España se descubren alrededor de 22.000 casos al año [38].

Se habla de cáncer cuando, al contrario de la situación normal, en la cual las células sanas del organismo se reproducen de forma lenta y controlada; éstas se ven afectadas por mutaciones del material genético con el paso de tiempo, derivando en un crecimiento celular anormal y finalmente en la formación de un tumor, que adquiere la capacidad de invadir tejidos cercanos (infiltración) y de proliferar en otras partes del organismo (metástasis). Esto es lo que se conoce como tumor canceroso o maligno, si bien cabe mencionar que también puede darse el caso de que el crecimiento celular anormal sea lento y similar al de las células originales, y el tumor que se origine no suponga ningún tipo de peligro para la salud, que sea un tumor benigno.

En el caso de que las células mutadas sean las de la mama, bien las de las glándulas que producen la leche mamaria (cáncer carcinal) o bien las de los conductos que la transportan (cáncer ductal) [39], es cuando estamos ante el cáncer de mama. Este tipo de cáncer afecta mayoritariamente a mujeres, de las cuales 1 de cada 8 se verá afectada por cáncer de mama a lo largo de su vida, y suponiendo esta enfermedad el 25% del total de diagnósticos de todos los tipos de cáncer anuales[40]. Aunque a priori estos datos pueden parecer insalvables, los referidos a las tasas de supervivencia son también muy altos. Un diagnóstico temprano del tumor maligno en sus primeras etapas, y su consecuente tratamiento, hará que el problema se resuelva en un 80-90% de los casos. Por tanto, lograr un diagnostico lo más rápido posible y veraz resulta un objetivo claro y necesario en el que trabajar.

A la hora de detectar el cáncer de mama en etapas tempranas, los médicos buscan pequeños tumores o algún otro indicio que denote la presencia de la enfermedad. Una de las lesiones más frecuentes junto con las masas tumorales son los grupos de microcalcificaciones, pequeñas acumulaciones de calcio cuya presencia está relacionada con una elevada actividad celular, como la que se produce en el caso de la aparición de un tumor. Se sabe que entre un 30% y un 50% de los pacientes afectados de cáncer de mama presentan dichas microcalcificaciones en sus mamografías, y entre un 60% y un 80% lo hacen en exámenes histológicos [41]. Por ello, aunque su presencia no implica de forma estricta la existencia de cáncer, sí que es un buen indicador de ello, y por lo tanto es útil para detecciones tempranas de tumores.

Para la detección existen diversos métodos (exámenes físicos, biopsias para pruebas de laboratorio como la FNB *(Fine Needle Biopsy)*, pruebas genéticas) pero si hay uno que resalta sobre los demás son las pruebas de imágenes médicas (rayos X, resonancia magnética, ultrasonidos y tomografías), y en particular las mamografías [41].

El extendido uso de la mamografía se debe, por un lado, a que es una técnica mínimamente invasiva, y por otro, a la gran fiabilidad de esta prueba. Las mamografías no solo se emplean para la detección del tumor, sino también para su evaluación y seguimiento. Para generar mamografías un equipo de rayos X proyecta dicha radiación, que atraviesa el tejido de la mama en mayor o menor cantidad según la densidad del tejido, obteniéndose una imagen en escala de grises que permite identificar las zonas dañadas. Lo más común hasta el momento era hacerlo de manera analógica, pero en los últimos años se han comenzado a sustituir los equipos analógicos por digitales, lo cual ha llevado a enormes mejoras en la calidad

de las imágenes, y consecuentemente en la calidad de la detección, ya que las mamografías digitales pueden ser procesadas fácilmente para eliminar el ruido y los artefactos presentes, siendo así más fáciles de interpretar.

La interpretación de las mamografías es precisamente el mayor reto con el que un radiólogo se puede encontrar hoy en día. Debido al ruido y a la similitud de algunos tejidos, así como a las múltiples morfologías y variabilidad de las lesiones muchas veces resulta complicado la emisión de un diagnostico cien por cien fiable.

La consecución de imagines más claras, y en definitiva de resultados más precisos es la motivación de muchos estudios que se están llevando a cabo en los últimos años. Así, la introducción de técnicas automáticas de detección, segmentación, localización y clasificación, entre otras, son de gran ayuda para la toma de decisiones por parte del profesional. Estos sistemas pueden, en un futuro, llegar a realizar la labor de identificación de lesiones de forma íntegra, si bien está comprobado que por ahora funcionan como un complemento o como un segundo "experto" que le ayude a realizar un diagnóstico.

En concreto en este trabajo se efectúa un estudio de todos aquellos sistemas propuestos en los últimos años que emplean métodos de aprendizaje profundo para tareas relacionadas con la mejora del diagnóstico del cáncer de mama, como pueden ser la selección de características de forma automática, la detección de anormalidades y su clasificación en malignas o benignas, o la detección y localización de las lesiones. Además, partiendo de esta revisión del estado del arte, se propone una metodología para la detección automática de lesiones en las mamas, que será explicada en la siguiente Sección 4.3.

4.2. Estado del arte del Aprendizaje Profundo en imagen del cáncer de mama

Un radiólogo decide sobre la malignidad o benignidad de un tumor basándose en una serie de características específicas que encuentra en la lesión. Por ello, a la hora de automatizar este proceso se requiere de un sistema que tenga definidos estos descriptores de los tumores. Los sistemas que realizan esta labor de toma de decisiones son los CADe y los CADx, para la ayuda a la detección y para la ayuda al diagnóstico, respectivamente.

Estos sistemas funcionan a partir de unos datos que son procesados por el sistema, basado en algún tipo de modelo. En los últimos años se ha demostrado que el Aprendizaje Profundo tiene un gran potencial para ser este modelo de base. Lo más común desde los inicios de su aplicación al análisis de imágenes medicas era emplear como datos imágenes, en concreto mamografías, y CNNs como modelos para analizarlas. De hecho, el artículo que data de 1996 mencionado anteriormente lo confirma, ya que proponen una CNN para la clasificación de ROIs de mamografías en tejido normal o masas [37].

La mamografía es sin duda la modalidad de imagen mas empleada para estas tareas en el cáncer de mama, debido a que es el estudio más comúnmente realizado y por lo tanto del cual se disponen más datos, sin anotar, pero sobre todo anotados, lo cual es algo muy útil para entrenar los modelos de aprendizaje profundo. A pesar de que la mayoría de artículos revisados emplean mamografías, sí que se

encuentran otros que proponen usar otras modalidades de imagen, como pueden ser los US [42] [54] [60] , los MRI [47], y los volúmenes de TS [50] [57] [58].

Los sistemas de ayuda a la decisión por ordenador CAD (para generalizar y no hablar solo de CADx o de CADe) suelen realizar las tareas en tres pasos. El esquema más general empieza por una etapa de obtención, tratamiento y adecuación de los datos, que puede ser manual o estar automatizada. A continuación, está una etapa de detección y localización de lesiones, que puede estar más o menos automatizada también, y finalmente se encuentra la etapa de clasificación. Para las distintas tareas se requieren distintos tipos de modelos, y los trabajos que aquí se analizan se centran o bien en alguna de ellas o en el proceso global.

Para la segunda etapa, que es donde más peso tienen los modelos profundos, en los últimos años sigue habiendo muchos trabajos que proponen a las CNNs para las etapas de extracción de características, detección y localización de las lesiones [44] [46] [47] [48] [51] [52] [56], si bien existen otras técnicas de aprendizaje profundo con las que también se está experimentando para el mismo fin, como las ADN [42], variaciones de la Máquina de Bolztmann como DBN [64] y RBM [60], y los autocodificadores [53] [63] y todas ellas obtienen también buenos resultados, a pesar de que se ha demostrado que el aprendizaje de este tipo de redes es algo más complicado y costoso [42].

Así, volviendo a las CNNs, la forma de operar típica para la detección de anormalidades, bien sea en una mamografía o en otro tipo de imagen médica, se divide en una serie de pasos. Para empezar se generan una serie de candidatos que pueden ser distintos tipos de anormalidades, generalmente masas o microcalcificaciones, definidos por sus coordenadas, y a continuación son todos ellos pasados por un extractor de características que han sido definidas y extraídas manualmente en una fase previa [50], quedándose el sistema solo con algunos de ellos que serán salida de esta red y entrada de la siguiente capa, la encargada de la clasificación (si el sistema tiene como propósito clasificar y no solo detectar y localizar).

Aunque sí que es cierto que los sistemas suelen emplear para el diagnóstico de cáncer el criterio de presencia en la imagen de microcalcificaciones y/o masas, cada vez se encuentran más trabajos que proponen otros criterios para predecir el cáncer de mama, como pueden ser la densidad del tejido parenquimal [44], que está directamente asociada con el desarrollo de cáncer de mama, la cantidad relativa de tejido radiodenso o la heterogeneidad del pecho según sus características de textura [53] [63]. Aun así, no está todo hecho todavía en los sistemas que buscan detectar y clasificar lesiones y buscar este fin sigue siendo de interés, por ejemplo porque microcalcificaciones son a menudo pequeñas y un radiólogo puede omitirlas por error, por ello es importante seguir trabajando en este campo [55].

En cuanto a extracción de características, cuando es manual, tiene que ser realizada a partir de un conocimiento teórico y verificado, es decir, tiene que trabajar un radiólogo experto, o más de uno, en elegir estas características y en anotar las imágenes, a nivel de pixel o de regiones, para que pueda usarse como base del aprendizaje. Algunos ejemplos de características empleadas son histogramas de valores de intensidad, características de morfología y descriptores de textura la forma, el tipo de frontera o la densidad [54], aunque muchas veces es

útil añadir otro tipo de características como de ubicación y de contexto, e incluso no inherentes a la propia lesión, como pueden ser datos del paciente [55]. Pero esta extracción de características está ligada a la necesidad de disponer de un conjunto de datos anotados por un experto muy grande. Como es lógico pensar, esto es algo muy complicado porque para hacerlo de forma correcta se necesitaría a más de un radiólogo que evaluase y etiquetase las imágenes, además de tener que lograr un acuerdo de confidencialidad por parte de los médicos y de los pacientes, que normalmente se oponen a la compartición de los datos [59].

Por ello, por lo tedioso que es este proceso de anotación de lesiones y extracción de características de forma manual, se han comenzado a explorar otras alternativas que automaticen todavía más los CAD. Estas nuevas opciones se pueden dividir en dos grandes grupos: usar modelos más profundos y usar técnicas del conocido como Aprendizaje de Transferencia (TL o *Transfer Learning*).

Se comienza hablando acerca de los modelos de CNN profundas (DCNN), puesto que es el tipo de modelo que más adelante se propone en este trabajo para el análisis de mamografía. Estas redes están formadas por muchas más capas y por lo tanto tienen la capacidad de extraer conocimiento a niveles de extracción mayores, de manera mucho más profunda, como su propio nombre indica. Lo que proponen los autores de sistemas basados en estos modelos [49] [50] [54] [55] [57] [58] [59] es que a partir de las imágenes, sea la propia red la que aprenda las características, sin necesidad de anotaciones a nivel de pixel, es decir, sin necesidad de que un especialista marque las ROI por ejemplo, si no solo a nivel de imagen [54]. Para ello los sistemas generan los candidatos de lesión de forma automática y usan la localización para situar en ese punto el centro de la ROI, generando *patches* de estas lesiones, así como de las demás partes de la imagen [50] [55]. El empleo de *patches* o sub-imágenes aporta múltiples ventajas, pues cogen toda la información de los candidatos además de algo de información del contexto de su entorno, lo que es útil para obtener localizaciones más precisas y para la posterior clasificación [50]. Consecuentemente no requieren ningún tipo de información de localización espacial, y el añadírsela no mejora los de resultado [48]. Otros sistemas basados en DCNN emplean aproximaciones similares pero sí que requieren trabajo de anotación, aunque en menor medida, como se puede ver en [57] y [58] donde se pide a un experto que seleccione solo las ROIs que son de interés, las positivas, y el sistema se encarga de la detección de todas las demás.

Sin embargo, lo común es que requieran una mayor gran cantidad de datos etiquetados para el entrenamiento porque precisamente tienen que aprender de ellos de forma no supervisada o débilmente supervisada, lo que sigue siendo algo complicado a la hora de trabajar con imágenes médicas, por los motivos ya comentados. Aun así, el potencial de las DCNN está claro que es altísimo.

La segunda alternativa a las CNNs simples es usar el Aprendizaje de Transferencia, estrategia que cuenta con múltiples ventajas. Una de ellas es que no se requieren imágenes médicas etiquetadas para que los resultados sean buenos a la hora de clasificar dichas imágenes. Esto funciona porque las redes que funcionan por Aprendizaje de Transferencia son pre-entrenadas en otro conjunto de imágenes que sí que están etiquetadas, de donde aprenden las características, y luego se re-entrenan en la tarea de interés. Este pre-entrenamiento puede hacerse en imágenes de una tarea no médica [52]; en imágenes de otra modalidad médica, es decir,

entrenar la red para una amplia base de datos de mamografías y "transferir" el conocimiento a una pequeña de TS como proponen los autores de [58], construyendo así un sistema multitarea, y también con un subconjunto del conjunto de las propias imágenes que se quieren clasificar, lo que se conoce como *Self-Transfer Learning* o Aprendizaje de Autotransferencia [57] [58] [59]. Este último tipo de aprendizaje sería como tener un modelo de aprendizaje semisupervisado, estando disponible un pequeño conjunto de datos etiquetados (o de ROIs seleccionadas) y un gran conjunto de datos sin etiquetar, que aprenden de las etiquetadas y van pasando a formar parte de este primer grupo.

Finalmente, la etapa de clasificación se realizará en un último paso mediante otras técnicas, en la mayoría de los casos de aprendizaje de máquina y no específicas del Aprendizaje Profundo, como pueden ser los SVM [42] [44] [46] [60] o los *Random Forests* [45] [49] [55], si bien se encuentra algún trabajo donde la parte de la red que hace la clasificación sí que es un algoritmo propio de aprendizaje profundo, como una DNN para dar probabilidades [48]. En muchos trabajos se emplea el esquema CNN+SVM, para detección y clasificación respectivamente, y lo denominan R-CNN [23] [44] [46] [53].

En esta etapa de clasificador se pueden buscar distintos tipos de salidas. Lo tradicional es realizar una clasificación binaria de las lesiones detectadas, es decir, una clasificación en dos clases, maligno/benigno, cáncer/no cáncer, etc. pero otros autores están proponiendo clasificaciones alternativas, ligadas a buscar patrones de textura y densidad y no lesiones en las imágenes [44] [53]. También se encuentran clasificaciones asociadas al tejido de la mama (músculo pectoral, tejido fibroglandular, pezón, tejido general del pecho, que incluye al tejido graso y a la piel) [48], y por último clasificaciones realizadas según el estándar BI-RADS [45], que es un sistema elaborado por el *American College of Radiology* para estandarizar esta clasificación de lesiones de la mama permitiendo que pueda ser comprensible tanto para otros radiólogos de otros centros, como para lectores que no sean radiólogos ni especialistas. El BI-RADS clasifica los resultados y hallazgos de las mamografías en 7 categorías del 0 al 6, representando el 0 la necesidad de estudios adicionales, el 1 un resultado negativo, y los números del 2 al 6 como resultados que afirman la presencia de un tumor con mayor grado de malignidad en escala creciente [65].

Cabe destacar asimismo la finalidad de dos trabajos en concreto que se desmarcan bastante de estas tendencias. En [62] se busca detectar microcalcificaciones arteriales en el pecho o BACs para asociar su presencia con la posibilidad de que el paciente padezca de una enfermedad cardíaca, y en [56] se revisan de nuevo mamografías clasificadas como negativas para ver si se pueden encontrar indicios de posible desarrollo de un tumor en las mamas.

Una vez propuestos y construidos todos estos sistemas, es vital analizar su desempeño en términos de efectividad y precisión, además de comparar sus resultados con los de radiólogos especialistas y frente a sistemas tradicionales CAD más antiguos (menos automatizados, con extracción manual de características por ejemplo). Lo habitual es que por un lado, los sistemas con características automáticas proporcionen resultados parecidos a los antiguos, pero si se combinan con algunas características extraídas manualmente y siempre escogidas con criterio, apuntando a los puntos débiles de la red neuronal, los resultados suelen

mejorar bastante [46] [55], y de hecho en muchas situaciones es una estrategia a seguir mucho más efectiva que el buscar mayores cantidades de muestras que añadir al conjunto de entrenamiento. Por otro lado, los sistemas propuestos mejoran la precisión respecto a la de los especialistas por separado, pero no respecto a la de los especialistas en media [55].

Aunque todavía no se hayan alcanzado resultados significativamente mejores que los humanos hay que seguir trabajando en la construcción y mejora de estas redes. Existen numerosos factores que influyen en gran medida en la calidad de los resultados, y suelen ser métodos de pre-procesado o post-procesado de las imágenes. El pre-procesado incluye distintas técnicas de mejora de la calidad de la imagen y de eliminación de ruido, y que si no se aplican hacen que la precisión de la red decaiga notablemente [49]. Pese a que suelen ser pasos que se aplican por separado, existe algún sistema que ya propone integrarlos como una primera parte del mismo, logrando así un pre-procesado automático, como se puede comprobar en [45], donde se proponen una serie de regiones ya habiendo segmentado, eliminado el fondo de la imagen y cortado la región automáticamente. Las distintas formas de pre-procesado de la imagen serán explicadas en detalle más adelante cuando se explique la metodología diseñada en este trabajo. Otra etapa previa a alimentar la red que se puede considerar de pre-procesado es la selección de las ROI, de la cual ya se ha hablado previamente.

También cabe tener en cuenta dentro del pre-procesado las estrategias de aumento de datos, muy usadas en imagen médica para lograr un conjunto de datos de mayor tamaño. El aumento de datos suele seguir dos líneas; una más clásica a partir de transformaciones geométricas, ya sean rotaciones, traslaciones, inversiones y cambios de escala [55] [56] [57] [58] [62] y otra asociada a los trabajos ya revisados que emplean *patches* para entrenar sus redes, recortando sub-imágenes que se pueden superponer, de cada una de las imágenes, de forma que se consiguen multitud de sub-imágenes con combinaciones de pixeles distintas. Unido al aumento de datos surge la necesidad de hacer que las clases a las que pertenecen los *patches* estén balanceadas, para lograr un buen entrenamiento de la red. Para lograr esto se muestrean aleatoriamente una cantidad de *patches* de la clase mayor en número, equiparando los conjuntos [57] o se mantiene la desigualdad pero en el entrenamiento se corrige mostrando los *patches* en menor cantidad un mayor número de veces [55]. Los *patches* también llevan asociados habitualmente técnicas de submuestreo para reducir su tamaño y hacerlos todos de igual dimensión o técnicas de post-procesado para reconstruir la imagen original a partir de ellos.

A modo de conclusión de esta sección, es importante resaltar dos cosas. Primero que aun no alcanzando resultados de precisión máxima, todos estos modelos están superando poco a poco el estado del arte con el que compiten, lo cual es una muy buena señal. Y segundo, que todos estos modelos no buscan en ningún caso la sustitución completa, si no el ayudarle, tanto liberándole de realizar algunas tareas que la máquina realiza mejor, como el asesorarle en algunos puntos del proceso, o como el proporcionar una segunda opinión a modo de "segundo profesional", lo que en cualquier caso es de gran ayuda para el radiólogo.

En la Tabla 2 se recogen los distintos trabajos mencionados en las líneas anteriores y se detalla, para cada uno de ellos y siempre que sea posible, la

información considerada como más relevante para cada uno de ellos, es decir, su finalidad, la modalidad de imagen empleada, el pre-procesado de los datos, la forma de obtener de obtener las características, el método empleado y los resultados obtenidos por los autores. Si bien en ningún caso estamos ante modelos que funcionen en un CAD en la vida real, todavía; los propios autores se encargan de comprobar la eficacia de sus sistemas, y proporcionan resultados empleando diversas métricas así como comparaciones con otros trabajos. Por ello, se van a describir cada una de las métricas que estos autores emplean, para su correcta compresión. La métrica más empleada es el área bajo la curva ROC (*Receiver Operating Characteristic*), conocida como AUC, que representa los verdaderos positivos (VP) frente a los falsos positivos (FP), por lo que es una medida de la calidad de la clasificación, siendo 1 su valor máximo y 0 su valor mínimo. Otros trabajos recogen simplemente la tasa de VP y de FP, y otros hablan de la especificidad, que se define como la capacidad de detectar aquello para lo que ha sido creado el clasificador (ratio de verdaderos negativos entre el total de negativos detectados), y de la sensibilidad, que es la habilidad para detectar los verdaderos positivos (ratio de verdaderos positivos entre el total de positivos detectados). En menor medida se emplean el índice de Dice, que compara la similitud entre dos muestras; el valor Kappa, para medir la concordancia de la clasificación entre dos sistemas, y la desviación estándar (STD).

Tabla 2: Resumen de los trabajos más actuales que emplean modelos de Aprendizaje Profundo para el análisis de mamografías. Para cada uno de ellos se detallan los componentes del sistema más relevantes, junto con la modalidad de imagen, la finalidad del sistema y los resultados obtenidos.

Referencia	Año	Modalidad imagen	Finalidad	BBDD	Pre-procesado	Obtención parámetros	Método	Resultados
[37]	1996	MG	Clasificación de ROIs en tejido normal / masas	168 mamografías	Mamografías digitalizadas Normalización de las imágenes ROIs seleccionadas manualmente y sub-muestreadas a 16x16 y 32x32	4 características de textura calculadas a partir de las imágenes	CNN	ROC=0.87
[42]	2012	MG US	Clasificación de lesiones cancerosas/no cancerosas	Conjunto de 739 MG y 2393 US de el Centro Médico Universitario de Chicago (UCMC)	Sub-muestreo de las ROIs a 256x256 Recorte a 140x140 Relleno de las imágenes si necesario	Extracción de características no supervisado (aprendizaje automático a partir de la imagen)	**ADN** + SVM	AUC (US) = 0.83 AUC (MG) = 0.71
[43]	2015	MG	Detección de masas	INbreast (410 imágenes) DDSM-BCRP (79 casos)		Extracción de características y clasificación Detección de candidatos automática por medio de una máscara a diferentes resoluciones	**R-CNN (CNN** + SVM)	Para INbreast: TP=0.96 at FPI= 1.2 TP of 0.94 at FPI= 0.3
[44]	2015	MG	Clasificación según la densidad del tejido parenquimal (4 clases)	Mamografías de 1157 mujeres	ROI seleccionada manualmente Redimensionado de las imágenes a 200x200 píxeles, filtrado, umbralización y normalización	Extracción automática de características	**CNN** + SVM	Precisión = 66.96% Eficacia (kappa=0.58) análoga a la de un experto (kappa=0.56-0.79)
[45]	2016	MG	Detección de anormalidades y clasificación en normales, benignas, malignas (puntuación BI-RADS 1-5)	850 mamografías	Pre-procesado automático para la segmentación del tejido de la mama	Propuesta de regiones candidatas automática y selección automática Aumento de datos	**DBN + R-CNN** + 2 Random Forests	Precisión del 0,775% AUC = 0.66
[46]	2016	MG	Clasificación de masas	BCDR-FM (1010 casos)	Segmentación de la ROI Aumento de datos por sobremuestreo Normalización con contraste local y contraste global	Imágenes etiquetadas por experto Conjunto de 17 características extraídas manualmente + histograma para comparación Aprendizaje supervisado de características	**CNN** + SVM	ROC(caract. automáticas)= 0,822 ROC(caract. manuales+automáticas)= 0,826

Referencia	Año	Modalidad imagen	Finalidad	BBDD	Pre-procesado	Obtención parámetros	Método	Resultados
[47]	2016	MRI	Clasificación de lesiones en malignas y benignas Clasificación de distintos tipos de carcinomas	325 MRIs		Segmentación semiautomática por el algoritmo *Multiseed Smart Opening* 5 caract. morfológicas y 6 dinámicas obtenidas automáticamente	**CNN +** *Random Forest*	AUC=0.8543 AUC(DCI)=0.7924 AUC(IDC)=0.8688 AUC(ILC)=0.8650
[48]	2016	MG	Clasificación de píxeles en 4 categorías de tejidos del pecho	40 mamografías MLO	Umbralización para eliminar píxeles de fondo Post-procesado: Sub-muestreo de las salidas reducidas un factor de 8 para juntar los *patches*	Segmentación manual y etiquetado por un experto Generación de 800.000 *patches* Aprendizaje automático de características	**DNN**	Dice= 0.85 - 0.56 (Según tejido)
[49]	2016	MG	Segmentación automática de masas	INbreast (56 casos con 116 masas) DDSM-BCRP (158 imágenes)	Imágenes redimensionadas a 40 x 40 píxeles por interpolación y pre-procesadas con la técnica de Ball y Bruce	ROI obtenida manualmente Características aprendidas automáticamente	**DCNN** + CRM/SSVM	Dice (CRF) = 0.93 Dice (SSVM) = 0,95 Dice (NO pre-procesado)= 0,85
[50]	2016	TS	Detección y reconocimiento de cáncer (masas y distorsiones arquitectónicas positivas/negativas)	1864 lesiones sospechosas de MG 2D y 339 lesiones de volúmenes DBT.	*Patches* redimensionados a 256x256 por interpolación bilineal Brillo de los patches re-escalado Entradas pasadas de RGB a escala de grises	Generador de candidatos automático y de *patches* alrededor de cada uno que cogen info. de contexto evitando la extracción manual de características	**DCNN**	Sensibilidad(ROIs sospechosas)= 0.893 Sensibilidad(ROIs malignas)= 0.930 Precisión = 0.8640
[51]	2016	MG	Localización y clasificación de masas (positivas/negativas)	DDSM (10363 imágenes) MIAS (322 imágenes)	CXRs y mamografías redimensionados a 500x500	Modelo débilmente supervisado (información solo a nivel de imagen, ROIs marcadas)	**STL + CNN**	Localización mejora en un 242% y clasificación en un 6% en comparación con otros métodos
[52]	2016	MG	Clasificación de lesiones en malignas y benignas	607 FFMDs con 219 lesiones en total		Aprendizaje de Transferencia con una CNN pre-entrenada en tarea no médica y luego entrenada con las mamografías	**CNN**	AUC(TL)=0.81 AUC(TL+caract manuales)=0.86

Referencia	Año	Modalidad imagen	Finalidad	BBDD	Pre-procesado	Obtención parámetros	Método	Resultados
[53]	2016	MG	Clasificación y segmentación del pecho por puntuación de densidad MD Clasificación de tejidos por puntuación de texturas MT	Dataset MD, MT y *Dutch Breast Cancer Screening dataset* (493, 668 y 1576 imágenes)	Redimensionamiento de las imágenes	Segmentación de mamografías en fondo/musculo/tejido mama (ROI) Generación automática de *patches* de 24x24 y aprendizaje de características a partir de datos sin anotar	**CNN + CSAE**	AUC=0.59
[54]	2016	US MG	Detección y descripción semántica de lesiones (Masas) Sistema multitarea	974 MG de DDSM 646 MG propias 408 US propias		Generación automática de las ROI con cajas rectangulares y extracción automática de características Uso de distintos descriptores (forma, margen, orientación, fronteras, etc.)	**R-CNN**	Precisión(DDSM)=0.82-0.77 Precisión(US)= 0.82-0.78 Precisión(MG)= 0.88-0.84
[55]	2016	MG	Detección y clasificación de masas y microcalcificaciones en positivo/negativo	45.000 imágenes	Aumento de datos mediante transformaciones geométricas Relleno de imágenes con 0's si necesario Eliminación de ejemplos anotados del set de entrenamiento	Detección de candidatos y generación de ROIs automático con *patches* de 250x250 para la generación de características automáticas Conjunto de 74 características obtenidas manualmente	**DCNN +** *Random Forests*	AUC(sin aumento datos) = 0.875 AUC(aumento datos) = 0.929 AUC(aumento+caract. manuales) = 0.941
[56]	2016	MG	Desarrollo futuro de cáncer de pecho a partir de mamografías negativas	270 clasificadas como negativas			**1CNN +** MLP	Precisión = 71.4%
[57]	2016	TS	Detección de microcalcificaciones y clasificación en verdaderas/falsas	64 casos de DBT con microcalcificaciones (64 vistas CC + 63 vistas MLO)	Pre-procesado de volúmenes para obtener las imágenes Filtrado del ruido de alta frecuencia Transformaciones geométricas para aumentar el conjunto de datos Balanceo de datos	Microcalcificaciones verdaderas anotadas manualmente y ROI de 16x16 en el centro de cada una Eliminación de FP por otro CAD Detección por umbralización iterativa y crecimiento de regiones, agrupamiento, y reducción de FPs	**DCNN**	AUC=0.93 (Comparado con una CNN con AUC=0,89)

Referencia	Año	Modalidad imagen	Finalidad	BBDD	Pre-procesado	Obtención parámetros	Método	Resultados
[58]	2016	TS	Aprendizaje de Transferencia para clasificación de masas en TS a partir de MG	2282 mamografías anotadas 324 DBTs	Aumento de datos de ROIs Normalización de ROIs	Aprendizaje de transferencia ROIs VP segmentadas manualmente y FP eliminadas con otro CAD	**DCNN**	AUC=0.90 (=0.81 sin TL) Sensibilidad=91% (=83% sin_TL)
[59]	2016	MG	Diagnóstico de cáncer	1874 pares de mamografías (un corte CC y un corte MLO)	Segmentación de subregiones, extracción de características bilaterales, selección de características, clasificación, etc. Aumento de datos por transformaciones geométricas	Aprendizaje semi-supervisado (Pocas ROIs etiquetadas + muchas ROIs sin etiquetar) 21 características anotadas manualmente y otras obtenidas a posteriori	**DCNN**	AUC=0.8818 Precisión=0.8243
[60]	2016	US	Selección automática de características Clasificación de tumores malignos / benignos	227 imágenes SWE	Sub-muestreo de las imágenes a un tamaño y resolución fijos	Selección de características automática (sin segmentación ni datos anotados) Extracción manual de 286 características simples para comparación	**PGBM + RBM** + SVM	Precisión=93.4% Sensibilidad=88.6% Especificidad=97.1% AUC=0.947
[61]	2017	MG	Discriminación de quistes aislados benignos y masas malignas	1000 imágenes con masas malignas y 600 imágenes con quistes	Imágenes transformadas logarítmicamente, invertidas y segmentadas Aumento de datos a nivel de tejido	Extracción manual de *patches* de 260x260 Extracción de 5 características manuales. Combinación posterior con características automáticas. Red pre-entrenada con *patches* de masas normales	**DCNN**	AUC = 0.80 AUC=0.87 si se usan solo lesiones mayores de 20 mm
[62]	2017	MG	Detección de BACs (Calcificaciones arteriales del pecho)	840 FFDMs de 210 casos de distintos centros		Subimágenes de 95x95 pixeles Aumento de datos (Rotacion de 90º,180º y 270º y flippings)	**DCNN**	Identificación de BACs similar a la de lectores humanos
[63]	2017	MG	Clasificación por puntuaciones de MD (tejido denso/tejido graso/fondo) y de MT	Mini-MIAS	Eliminación ruido por filtrado Supresión de artefactos radiopacos por umbralización Segmentación del fondo	Obtención de ROIs por crecimiento de regiones	**Autocodificador** + Clasificador *softmax*	Precisión=98,5%

4.3. Diseño de una metodología para el análisis de mamografías

En este último punto del trabajo, y tras haber estudiado en profundidad las bases del Aprendizaje Profundo, así como los algoritmos y modelos que se están aplicando actualmente para el análisis de imágenes de la mama, se va a proponer una nueva metodología para analizar mamografías. Así, se comenzará por explicar la base de datos de la que se parte, y a continuación se justificará la elección del modelo propuesto y se explicarán sus distintas partes y componentes.

4.3.1. Base de datos

En primer lugar se describe la base de datos sobre la cual se entrenaría, se validaría y se probaría la metodología diseñada. Existen distintas bases de datos de mamografías públicas y conocidas como son la Mini-MIAS, la DDSM, la B-SCREEN, la AMDI y la IRMA [66], de las cuales muchas de ellas se han usado en los trabajos recogidos en la Tabla 2.

En este trabajo se propone utilizar la base de datos DDSM o *Digital Database for Screening Mammography* [67] [68], por ser la única base de datos de mamografías digitales públicamente disponible. Contiene un total 2620 exámenes de distintos pacientes, cada uno de ellos formado por cuatro vistas, una cráneo-caudal (CC) y otra medio-lateral oblicua (MLO), para cada pecho. Así, en total contiene 10.480 mamografías, de tamaño variable, en torno a 3000x5000 píxeles, con una resolución entre 42 y 100 μm/píxel. Al ser las imágenes mamografías digitales, al trabajar con ellas se tiene la ventaja de que no contienen el ruido causado por la etapa de digitalización de las imágenes analógicas, presente en muchos *datasets*, y consecuentemente asegura una mejor calidad de estas mamografías. Además, las imágenes ya han sido pre-procesadas, cortadas para eliminar así el fondo y procesadas para oscurecer los píxeles que contenían los identificadores del paciente.

Además de las imágenes mamográficas en sí, la BBDD proporciona información adicional, tanto acerca del paciente al que pertenece cada mamografía (sexo, edad, raza) como una descripción relativa al análisis de las imágenes y realizada por un radiólogo experto; por ejemplo el número y tipo de anomalías presentes en la imagen, su localización, y el grado de fiabilidad del diagnóstico. En función de esta información las mamografías están ya clasificadas en cuatro grupos, por exámenes, de menor a mayor grado de severidad: exámenes con tejido normal (el examen resultó normal la primera vez así como al repetirlo una segunda vez), exámenes benignos sin un segundo examen (se calificaron como benignas la primera vez y no se vio necesidad de pedirle al paciente que volviera para un segundo examen), exámenes benignos (si el paciente tuvo que volver a ser llamado para un segundo examen), exámenes cancerosos (si en alguna de las imágenes se encontró prueba histológica de cáncer). Como esta clasificación es a nivel de examen se propone la clasificación a nivel de imagen descrita en el Trabajo de Fin de Grado [69], que divide las mamografías en los cinco siguientes grupos: imágenes con microcalcificaciones benignas, imágenes con masas benignas, imágenes con microcalcificaciones cancerosas, imágenes con masas cancerosas, e imágenes con tejido normal.

La descripción del conjunto de imágenes que se emplean para entrenar la red es algo esencial pues según sean de un tipo o de otro la arquitectura tendrá que

diseñarse de diferente manera, teniendo en cuenta sus características. Los resultados que se obtienen a la hora de implementar cualquier esquema profundo dependerán fuertemente de los datos con los que la red sea alimentada, no solo de su contenido sino también de su tamaño, siendo siempre mejor la clasificación cuanto mayor sea el número de imágenes de la base de datos. A la hora de trabajar con una serie de imágenes hay que tener en cuenta aspectos como la presencia de ruido y cómo se puede eliminar, si se dispone de información adicional de las imágenes, o si las clases en las que se quieren clasificar las imágenes se encuentran en igual proporción en el conjunto de datos de entrenamiento. Esto último junto con las particiones de los conjuntos de prueba, validación y entrenamiento siempre de forma aleatoria nos da la seguridad de no introducir sesgos en el método propuesto.

4.3.2. Pre-procesado y adecuación de los datos

Como ya se ha introducido previamente, cualquier sistema de análisis de imagen médica se compone de una serie de partes, que como se puede ver en el esquema de la Figura 6 son: El pre-procesado de las imágenes, que incluye una etapa de filtrado y mejora, y otra etapa de adecuación de las imágenes al modelo, la elección y construcción de la red para la tarea de detección/localización, y la elección del modelo para la clasificación final de las imágenes. Cada una de ellas se detalla a continuación.

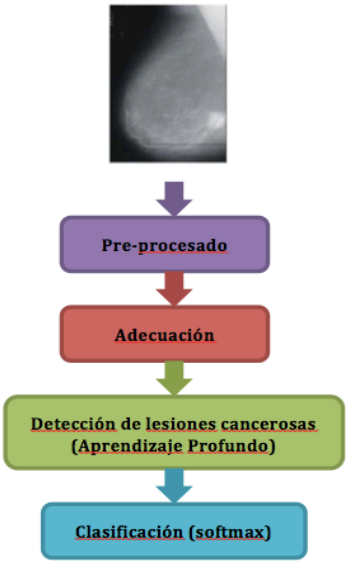


Figura 6. Secuencia de pasos del sistema propuesto

Mejora de la calidad de las imágenes

Esta primera parte del modelo consiste en la mejora de lo que serán las entradas de la red, y es un paso fundamental para que el entrenamiento de la red sea bueno y consecuentemente los resultados obtenidos [70]. Como es lógico pensar, está estrechamente relacionado al tipo de imagen que tengamos y sus características. Así, aunque las imágenes de la base de datos DDSM ya tienen algún tipo de pre-procesado realizado, como se puede ver en la Figura 7 es conveniente aplicarle algunas transformaciones para mejorarlas.

En la literatura los métodos más empleados son los de realce de imágenes, que pueden ser o bien de realce en el dominio espacial o bien de realce basado en características de la imagen.

Los primeros realizan las modificaciones a nivel de píxel, y los más populares son la modificación del histograma y la aplicación de filtros lineales espaciales. Existen otras técnicas más novedosas y especialmente útiles para el realce de características, que aunque son a nivel de píxel tienen en cuenta los valores de sus vecinos; es decir, efectúan las modificaciones a nivel de regiones.

A diferencia, los segundos modifican las imágenes en función de sus características, relacionadas o no con la organización espacial de la imagen, y no en las de cada píxel. Ejemplos de ello son la transformada wavelet y realce por lógica difusa.

En este trabajo se propone un pre-procesado dividido en tres pasos para eliminar así todos los posibles tipos de ruido que se pueden encontrar en la mamografía, señalados en la Figura 7. A partir de la mamografía original primero se le elimina el ruido y se le aplica un realce, y a continuación se suprimen los artefactos radiopacos y otros elementos no deseados, y se separará el pecho del fondo. Se propone seguir estos pasos en este orden, si bien se puede seguir la estrategia de la manera inversa[71].

Primero, para eliminar el ruido e incrementar el contraste de las imágenes a su vez se proponen tres métodos diferentes; la ecualización del histograma, la descomposición de la imagen mediante paquetes wavelet, y la aplicación de un filtro de mediana. El primero es muy sencillo y efectivo y potencia el contraste de los niveles de gris, permitiendo así una mejor diferenciación de los tejidos de la mamografía. Una alternativa posible a esta técnica sería aplicar una ecualización adaptativa del histograma con contraste limitado o CLAHE, que reduce de forma eficaz el ruido en regiones homogéneas y surgió específicamente para aplicaciones médicas [71]. A la hora de realizar el CLAHE es frecuente aplicar también a la imagen un filtrado de Weiner, para así lograr una mejora de contraste además de la eliminación de ruido [72]. El segundo consiste en aplicarle a la imagen una serie de filtros de descomposición, una umbralización, y finalmente volver a aplicarle los filtros para la recomposición de la imagen, logrando así eliminar el ruido aleatorio de distintas frecuencias. El tercero y último consiste básicamente en reemplazar el valor de cada pixel por el valor mediana de sus vecinos, y con ello se logra eliminar el ruido de las imágenes mamográficas también de una forma eficaz.

A continuación, para eliminar ciertos artefactos que pueden estar presentes en las imágenes, como las etiquetas, se propone usar una serie de operaciones morfológicas unidas a la operación de umbralización. Para ello, habría que inspeccionar las mamografías para ver cúal es el valor de umbral más adecuado, y una vez elegido se pasarían las imágenes de escala de grises a binario. El proceso de aplicación de las distintas operaciones morfológicas consiste básicamente en identificar las diferentes regiones de la imagen y quedarnos solo con la más grande, que se corresponderá con el pecho. De esta forma también se logra una segmentación del fondo de la imagen. Para volver a la imagen en escala de grises basta con multiplicar el resultado anterior por la imagen original.

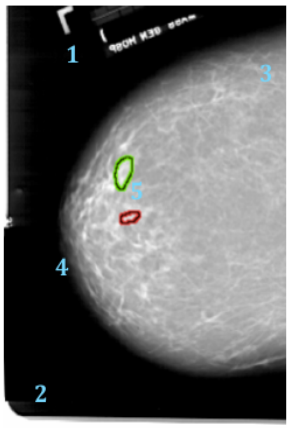


Figura 7: Mamografía original sin pre-procesar extraida de la DDSM [73]. 1) Ruido de baja frecuencia. 2) Fondo. 3) Músculo pectoral. 4) Perfil del pecho. 5) Lesiones.

Adecuación de las imágenes

Tal y como se explica en el apartado siguiente, la red neuronal que se pretende emplear en este sistema es una red neuronal profunda, y por lo tanto compuesta de muchas capas. Debido a las muchas capas por las que pasan las imágenes y para aliviar efectos de cargas computacionales altas, se opta por adecuar las mamografías a los requisitos de la red.

Primero, como no todas las imágenes son del mismo tamaño, sino que están entorno a los 3000x5000 píxeles, se recortan todas ellas para que sean del mismo tamaño, seleccionando un área entorno a la región de interés, y luego se subdividen en *patches* de 32x32 píxeles. En este caso se escoge este tamaño de *patch* para adoptar un esquema lo más similar posible a *DenseNet*, pero podrían usar otros tamaños mayores (64, 96, 224, 512 también se usan en otros trabajos) siempre que fueran divisibles por 2. Subdividiendo cada imagen en otras más pequeñas se logra aumentar el conjunto de datos y reducir el tamaño de las entradas a la red sin perder las características que se encuentra a escala muy pequeña, lo que sucedería si se realizase un sub-muestreo de la imagen. Además, para obtener variabilidad en las entradas y evitar el *overfitting*, se aplica una estrategia típica de aumento de datos mediante transformaciones de las imágenes, creando así muchas más imágenes artificiales. En el modelo en el que está basado esta red [23] se propone un aumento de datos siguiendo las nociones de [22], si bien estas redes se prueban en conjuntos de imágenes no médicas, de gran tamaño, y en algunos de los casos con imágenes en color, por lo que no se contempla como la aproximación más adecuada para este caso. Por otro lado, en [69] se plantea extraer aleatoriamente zonas de 24x24 píxeles de las imágenes, y aplicarles reflexión horizontal. Esto podría ser una opción, aunque resulta más óptima la de rotar las imágenes 90, 180 y 270 grados y voltearlas que se propone en distintos trabajos acerca de la adecuación de las imágenes mamográficas de la DDSM [75] [76], obteniendo así de una imagen un total de ocho muestras, y sin necesidad de disminuir más el tamaño de los datos de entrada, pues se confía en el poder de la red profunda propuesta para extraer todas las características relevantes.

El conjunto de imágenes que se tienen se dividirá de forma aleatoria para obtener los conjuntos de entrenamiento y de prueba de la red, y a su vez dentro del conjunto de entrenamiento se efectuará otra división aleatoria entre entrenamiento y validación, en una proporción 90:10, respectivamente. En ambos casos se

asegurará que haya un balance entre los distintos tipos de imágenes establecidos para cada conjunto.

4.3.3. Diseño de la Red Neuronal Convolucional Profunda

Como se ha podido ver previamente, las CNNs presentan muy buenos resultados a la hora de trabajar en la clasificación de lesiones de imágenes médicas, tanto en tiempo como en precisión. Sin embargo, tienen un problema principal; no son capaces de localizar las lesiones por sí solas de una forma suficientemente precisa. La explicación para esto es que no existen BBDD con información a nivel de pixel, si no solo a nivel de imagen, es decir, lo normal es que las mamografías contengan información de si presentan algún tumor o no, del tipo de tumor, etc. pero no que tengan localizadas y señaladas todas y cada una de las lesiones presentes, pues tener esto sería una tarea imposible.

Por ello, últimamente se está investigando acerca de modelos más profundos, como pueden ser las redes *Inception* [20] [21], *Highway Networks* [74], *ResNet* [22], para lograr características más ocultas de las imágenes. La idea básica de su funcionamiento es que, a más capas, mayor capacidad de captar características más abstractas de las imágenes, obteniendo conjuntos de características más completos, y en última instancia clasificaciones mejores.

Tras estudiar la estructura de estos modelos, a pesar de su baja tasa de error, surgen algunos problemas, principalmente el enorme número de parámetros que introducen tantas capas con tantas conexiones, además de otros subyacentes como la perdida de información a lo largo de la red debido a su longitud. Con esa motivación a finales del año pasado aparece *DenseNet* [23], un modelo que confía en conectar cada una de las capas con todas sus siguientes por concatenación. Esto ayuda entre otras a incrementar la variabilidad de las entradas de las siguientes capas (reutilización de características), a mejorar el flujo de información entre capas, a la supervisión de la red, y a reducir el número de parámetros, para lo cual agrupa las capas en bloques densos.

A la vista de los buenos resultados que obtiene en distintos conjuntos de datos, y por el interés que suscita el novedoso esquema de la red, en este trabajo se diseña una CNN profunda cuya estructura está basada en la de DenseNet, y cuyo esquema general se muestra en la Figura 8.

En esta sección se explica primero la estructura de la red que extrae las características, detallando sus parámetros y capas, y a continuación la clasificación realizada de las imágenes. La adecuación de las imágenes (en el esquema "*Patches* 32x32") se ha tratado en las sección anterior.

Así, la fase de extracción de características y reducción de parámetros se compone por una capa convolucional inicial (CONV) y una superposición de bloques densos y capas de transición (TRANS). Las dos capas de transición están formadas por una capa convolucional (CONV) y una de agrupamiento (POOL), mientras que cada uno de los bloques densos agrupa a un conjunto de capas convolucionales todas ellas con el mismo tamaño de filtro.

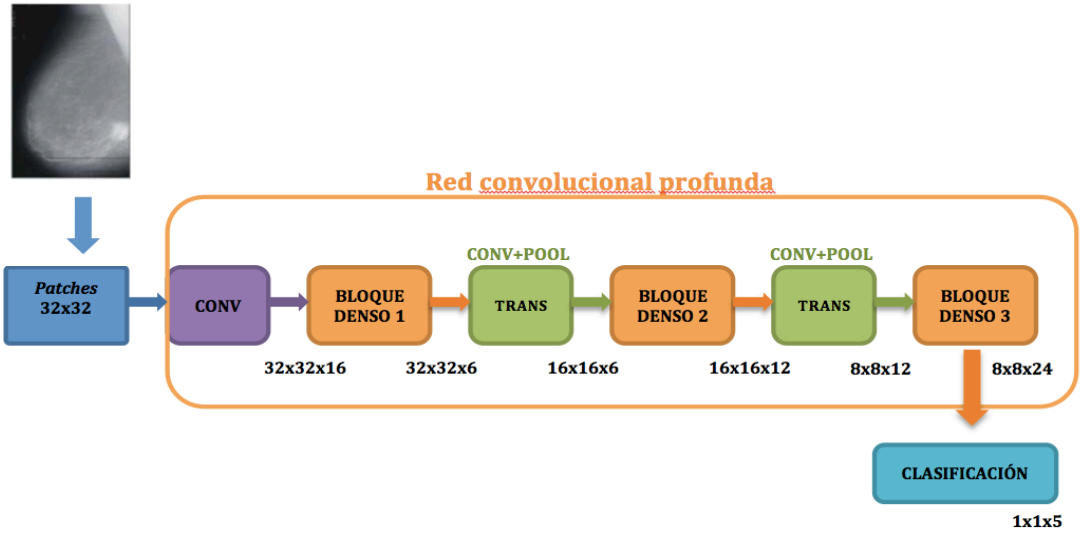


Figura 8. Estructura de la red convolucional profunda propuesta en este trabajo (capas y número de salidas por capa).

Capa		Salida	Estructura
CONV		32x32	5x5 CONV, paso=1, padding=2
BLOQUE DENSO 1		32x32	$\begin{pmatrix}1x1\ CONV\\3x3\ CONV\end{pmatrix}$ x 6
TRANS 1	CONV	32x32	1x1 CONV
	POOL	16x16	2x2 pooling de media, paso=2, padding= 0
BLOQUE DENSO 2		16x16	$\begin{pmatrix}1x1\ CONV\\3x3\ CONV\end{pmatrix}$ x 12
TRANS 2	CONV	16x16	1x1 CONV
	POOL	8x8	2x2 pooling de media, paso=2, padding= 0
BLOQUE DENSO 3		8x8	$\begin{pmatrix}1x1\ CONV\\3x3\ CONV\end{pmatrix}$ x 24
CLASIF	POOL	1x1	8x8 pooling de media global
	SOFTMAX		Clasificador lineal con 5 clases

Tabla 3: Resumen de características de cada capa que forman la red propuesta.

La primera capa en la red es una capa convolucional CONV, en la cual se aplica la operación de convolución para procesar las partes de la imagen de entrada extraídas en la etapa anterior. Aunque en [23] se propone usar un filtro de tamaño 7x7, se elige en este caso un filtro de tamaño 5x5, siguiendo lo establecido en [69] para esta primera fase, y dado que las buenas prácticas señalan emplear filtros de tamaño pequeño, pues se reduce así el número de pesos de la red.

En una operación de convolución el píxel de salida se calcula como una suma ponderada de los píxeles vecinos y que depende del núcleo de la convolución. De esta forma, variando los valores del núcleo se obtienen filtros personalizados para obtener distintas características de la imagen y como resultado queda una imagen filtrada en la que cada nuevo píxel contiene información de los píxeles vecinos en la relación definida mediante el núcleo de convolución.

Para prevenir que la red se haga demasiado profunda y mejorar la eficiencia de los parámetros, el número de filtros por viene limitado por la tasa de crecimiento de la red (k), y lo común es darle como valor un número entero

pequeño, pues aunque el número de mapas de características de salida no sea excesivamente grande, gracias a la conformación de la red en bloques densos se consiguen buenos resultados. Así, en este caso y siguiendo [23] se escoge un tamaño de k=32, y el número de filtros de esta primera capa se elige de 16, para cada canal, 3 en este caso. Por lo tanto, tras pasar esta primera capa las dimensiones de el volumen de entrada habrán pasado de ser de 32x32x3 a ser de 32x32x16. En cualquier caso, en las capas CONV se busca mantener la dimensión espacial del volumen de entrada a la salida. Por ello en este caso se emplea un *stride*=1 (S) y un *padding*=2 (P). Para calcular las dimensiones de la salida se usa la Ecuación (1) a continuación, donde O representa a la salida *(output),* I a la entrada *(input)* y F al tamaño del filtro *(filter*) [7] .

$$O = \frac{I - F + 2P}{S} + 1 \tag{1}$$

Además, esta primera capa convolución tiene aplicada tras ella una función de activación ReLU *(Rectified Linear Unit),* que es una función de activación basada en los elementos y que no varía las dimensiones del volumen.

Los pesos de una neurona definen los filtros aplicados en la convolución. Dado que en el número de pesos en una neurona corresponde al producto del tamaño del filtro y el número de canales de la entrada, en el caso de la capa CONV cada neurona tendrá un total de $5 \cdot 5 \cdot 3 = 75$ pesos distintos, estando estos pesos compartidos por todas las neuronas de esta capa.

Tras esta primera capa entra en funcionamiento lo que sería la parte de la red basada en la arquitectura de DenseNet. En esta arquitectura las capas se organizan por bloques, de forma que cada uno de los bloques agrupa a todas aquellas capas con el mismo número de filtros, siendo el número de filtros distinto entre bloques. En este caso se usan 6, 12 y 24 filtros para los bloques 1, 2 y 3, respectivamente. En cada bloque cada una de sus capas se conecta con el resto de capas hacia delante, teniendo por lo tanto múltiples conexiones directas, en total $L(L+1)/2$ conexiones, donde L representa el número de capas *(layers),* a diferencia de en una CNN tradicional donde se tendrían L conexiones. Así, para cada capa los mapas de características de las capas anteriores se utilizan como entrada, y su propio mapa de características se utilizará como entrada de todas las siguientes. Por lo tanto, la información obtenida en cada capa se va añadiendo a la información ya averiguada por concatenación, lo que hace que la red diferencie explícitamente entre la información nueva que se está añadiendo y la información que se preserva y que no tiene que añadir puesto que sería redundante. Este patrón de conectividad tan denso es lo que le da nombre la red y también a estos bloques, que se denominan “bloques densos”.

Así, para el primer bloque denso de la red, se tiene k=6 y por lo tanto solo 6 mapas de características por capa en este bloque, siendo notablemente más estrechos que lo habitual en una red profunda. En cada una de estas capas se tienen dos filtros CONV, primero uno de tamaño 1x1, que funciona a modo de cuello de botella, y a continuación otro de tamaño 3x3. Las capas convolucionales “cuello de botella” se introducen para reducir el número de mapas de características, dado que aunque cada capa produce solo k mapas de características de salida, normalmente tiene más entradas. Además, con ello se logra incrementar la

eficiencia computacional. La salida de este bloque es un mapa de características de dimensiones 32x32x6.

La conexión entre las L capas realizada entre cada uno de estos bloques no viene marcada únicamente por una función ReLU como pasaba para la primera cada del modelo. Tampoco emplea únicamente normalización por lotes (*Batch Normalization*, BN) como hacen otros modelos anteriores [22]. A diferencia implementa una transformación no lineal, que en la Ecuación (2) denominamos H_l, que es una función compuesta por normalización por lotes (BN) seguida de una unidad lineal rectificada (ReLU), y tras ella se aplica la convolución. De forma más específica, para este bloque denso, y para cada uno de los dos que siguen se tendría una estructura BN-ReLU-Conv(1x1)-BN-ReLU-Conv(3x3). Consecuentemente, el estado de un mapa de características dado viene descrito por:

$$X_l=H_l([x_0,x_1,\ldots,x_{l-1}]) \qquad (2)$$

donde los términos entre paréntesis denotan la susodicha concatenación entre las distintas capas del bloque. La operación de concatenación tiene el requisito de que los mapas de características no pueden ser tamaños diferentes, y de ahí la agrupación en bloques densos de las capas.

Tras haberse realizado el primer de extracción de características, se introduce su salida como entrada de una capa de transición TRANS, formada por una capa de convolución de 1x1, y a continuación se pasa por una capa de agrupación POOL, que es la que se encarga de reducir las dimensiones espaciales de la entrada. Aunque en la Sección 2 se decía que la estrategia de agrupación más habitual era la de *max_pooling,* para esta red se propone usar un *average_pooling* o de media, siguiendo las directrices de [23]. La cantidad de datos en este caso es reducida de manera que se ponen en común los pixeles de cierta zona de la imagen y se calcula la media de estos píxeles, que es lo que se devuelve como salida. Se ha escogido una agrupamiento con un tamaño de filtro de 2x2 , *stride*=2 y *padding*=0, de forma que se descartan exactamente el 75% de las activaciones en un volumen de entrada, porque se reducen las dimensiones a la mitad tanto en la altura como en la anchura de la imagen. Así, tras pasar el volumen por esta capa se obtendrá un volumen de salida de dimensiones 16x16x6.

El siguiente bloque denso y la siguiente capa de transición funcionan de forma análoga a las anteriores, pero en este caso el bloque denso está formado por 12 filtros, y por más neuronas, por lo tanto, y devuelve un mapa de características de tamaño 16x16x12, y la capa de transición se encarga de volver a reducir las dimensiones del volumen, resultando en un volumen de salida de la capa de transición 2 de tamaño 8x8x12, que será entrada de un último bloque denso. Este último bloque denso está formado por 24 capas agrupadas que comparten pesos, como hasta ahora, y por lo tanto su salida será un mapa de características de dimensiones 8x8x24.

Finalmente, y antes de introducir los datos resultantes en el clasificador para que efectúe su función de clasificador, se vuelve a pasar el volumen por una capa de agrupamiento para obtener un vector de dimensiones 1x1x(nº de clases). En este caso se emplea una función de agrupamiento de *pooling* de media global con tamaño de filtro 8x8, *stride*=1 y *padding*=0, calculado según la Ecuación (1).

Para elegir el clasificador, se ha pensado en el sistema propuesto como una ayuda para el profesional, una vez detectadas y localizadas de forma precisa todas y cada una de las lesiones, lo más útil es que el sistema proporcione un resultado que el medico pueda emplear intuitivamente. Por ello se propone emplear un clasificador sencillo con pocas clases como salida. Las aproximaciones más empleadas en la literatura, como se ha visto previamente, son los clasificadores SVM, *Random Forests*, o *Softmax*.

En este caso se emplea un clasificador *Softmax* [12], como última capa de la red, que se define como una red neuronal de dos capas, de las cuales una de ellas es una capa convolucional pre-entrenada, y por lo tanto supervisada. Una de las causas por la que se elige este clasificador es que devuelve probabilidades en lugar de márgenes, las cuales son mucho más fáciles de interpretar para un humano (no es así por ejemplo para el SVM). Para hacer esto, primero asigna unas puntuaciones de probabilidad logarítmicas no normalizadas para cada clase, y luego calcula las probabilidades normalizadas, de forma que, en este caso la de la clase incorrecta (más pequeña) y la de la correcta (más grande) deben de sumar uno. Es decir, como se tienen cinco tipos de clases de salida [normal, masa_benigna, masa_cancerosa, microcalcificación_benigna, microcalcificación_cancerosa], el clasificador devolverá un vector de tamaño fijo 1x1x5, como por ejemplo [0.9, 0.03, 0.01, 0.05, 0.01], donde se indica en este caso concreto que es muy improbable que este paciente tenga cáncer. Cabe destacar que para tomar esta decisión el clasificador habrá tenido en cuenta todos y cada uno de los mapas de características de la red, pues estos se han ido sumando condicionalmente los unos a los otros hasta llegar al final, como así lo propone DenseNet [23] .

5. Conclusiones y trabajos futuros

En el presente trabajo se han estudiado las bases del Aprendizaje Profundo así como los distintos modelos que se emplean, y las aplicaciones de estas técnicas. Con ello se ha aprendido en qué consiste este campo que recientemente ha logrado tanta atención, con el objetivo de despertar el interés en estas técnicas. El trabajo ha sido enfocado hacia el análisis de imágenes médicas, averiguando el sorprendente número de aplicaciones en distintos órganos y con distintos fines que estos modelos profundos están abarcando.

Partiendo de los buenos resultados que logran, se resalta una vez más la creciente importancia de los sistemas de ayuda a la detección y a la decisión en aplicaciones médicas que, sin reemplazar totalmente al profesional, le pueden servir de grandísima ayuda, facilitando su labor así como proporcionando mejores diagnósticos y demás servicios a los pacientes.

En concreto se ha querido centrar el estudio de estos algoritmos en imágenes mamográficas, al ver que era una de las líneas con más investigaciones. Para ello se ha hecho una profunda revisión bibliográfica de los artículos considerados más relevantes de los últimos años, con la que se ha logrado tener una completa visión de qué y cómo se está logrando automatizar la detección de lesiones y la clasificación de tumores en imágenes de las mamas. Se ha aprendido cómo se implementan los algoritmos y que no existe una regla para que funcionen, sino que es necesario un estudio heurístico; la importancia de estudiar el conjunto

de imágenes y de adecuarlas a la red para obtener buenos resultados; cómo leer los resultados, qué hacer para mejorarlos, etc.

Además, se han revisado los artículos relativos a los modelos de redes convolucionales profundas más utilizados en visión por ordenador. De la falta de aplicación de este tipo de redes al análisis de imagen médica, y entendidos los problemas con los que se enfrentan las redes neuronales más tradicionales ha surgido la motivación de este trabajo: diseñar una red convolucional profunda que se pueda emplear para el análisis de lesiones en mamografía. Este tipo de red aliviaría los problemas de amplias bases de datos anotadas requeridas, de médicos que tienen que señalar lesiones en las imágenes antes o localizar las regiones de interés como fase previa al entrenamiento del algoritmo, de costes computacionales altos, o de localizaciones imprecisas de las lesiones; todos ellos comentados a lo largo de este trabajo.

La red que se ha diseñado, aunque fundamentada en conceptos teóricos y en otros trabajos, es solo un esquema, y por ello sería un buen trabajo futuro y de gran interés poder implementarla y probarla. Dado que se dispone de la base de datos y del software necesario, se pretende dejarlo como una línea abierta para que sea continuada, de forma que a partir de este trabajo sea fáctible la implementación de esta red y la interpretación de los resultados, buscando como mejorarlos y comparándolos con los de otros trabajos.

Así, se puede concluir diciendo que los resultados obtenidos en este TFG son muy satisfactorios, pues aun no habiendo implementado la red diseñada, se ha estudiado de forma exhaustiva el aprendizaje profundo en imagen médica, un campo todavía muy nuevo y que suscita un gran interés, aprendiendo multitud de conceptos acerca de él y teniendo en estos momentos las bases necesarias para desarrollar más estudios en relación con la Inteligencia Artificial, tanto teóricos como prácticos.

6. Bibliografía

[1] Ian Goodfellow, Yoshua Bengio, Aaron Courville. *Deep Learning* (2016).

[2] Li Deng, Dong Yu. *Deep Learning: Methods and Applications* (2014).

[3] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, Michael S. Lew. *Deep Learning for visual understanding: A review* (2015).

[4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sánchez. *A Survey on Deep Learning in Medical Image Analysis* (2015).

[5] Jürgen Schmidhuber. *Deep Learning in Neural Networks: An Overview* (2016).

[6] Damián Jorge Matich. Redes Neuronales: *Conceptos Básicos y Aplicaciones* (2001).

[7] University of Standford. [En linea] Convolutional Neural Networks for Visual Recognition. Convolutional Neural Networks (http://cs231n.github.io/convolutional-networks/).

[8] A. Krizhevsky, I. Sutskever, G.E. *Hinton, Imagenet classification with deep convolutional neural networks.* En: Proceedings of the NIPS (2012).

[9] Amal Farag, Le Lu, Holger R. Roth, Ronald M. Summers. *A Bottom-Up Approach for Pancreas Segmentation Using Cascaded Superpixels and (Deep) Image Patch Labeling* (2016).

[10] Min Lin, Qiang Chen, Shuicheng Yan. *Network in network*. En: Proceedings of the ICLR (2013).

[11] University of Standford. [En linea] Convolutional Neural Networks for Visual Recognition [Repository of Stanford's CS231n GITHUB.]

[12] University of Standford. [En linea] Convolutional Neural Networks for Visual Recognition. Linear Classification (http://cs231n.github.io/linear-classify/#softmax).

[13] C. Szegedy, W. Liu, Y. Jia, et al. *Going deeper with convolutions*. En: Proceedings of the CVPR (2015).

[14] G.E. Hinton, N. Srivastava, A. Krizhevsky, et al. *Improving neural networks by preventing co-adaptation of feature detectors* (2012).

[15] N. Srivastava, G. Hinton, A. Krizhevsky, et al. *Dropout: a simple way to prevent neural networks from overfitting* (2014).

[16] L. Wan L, M. Zeiler, S. Zhang, et al. *Regularization of neural networks using dropconnect*. En: Proceedings of the ICML (2013).

[17] A. Krizhevsky, I. Sutskever, G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. En: Proceedings of the NIPS (2012).

[18] K. He, X. Zhang, S. Ren, et al. *Spatial pyramid pooling in deep convolutional networks for visual recognition*. En: Proceedings of the ECCV (2014).

[19] Karen Simonyan, Andrew Zisserman. *Very Deep Convolutional Networks for large-scale image recognition.* (2015)

[20] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. *Going Deeper with Convolutions* (2015).

[21] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Zbigniew Wojna, Jonathon Shlens. *Rethinking the Inception Architecture for Computer Vision* (2015).

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. *Deep Residual Learning for Image Recognition.*

[23] Gao Huang, Zhuang Liu, Kilian Q. Weinberger, Laurens van der Maaten. *Densely Connected Convolutional Networks* (2016).

[24] O. Russakovsky, J. Deng, H. Su, et al. *Imagenet large scale visual recognition challenge*. Int. J. Comput. Vis. 115 (3) (2015) p. 211–252.

[25] C. Szegedy, A. Toshev, D. Erhan. *Deep neural networks for object detection*. En:

Proceedings of the NIPS (2013).

[26] B. Zhou, V. Jagadeesh, R. Piramuthu. *ConceptLearner: discovering visual concepts from weakly labeled image collections.* En: Proceedings of the CVPR (2015)

[27] X. Liang, S. Liu, Y. Wei, et al. *Towards computational baby learning: a weakly-supervised approach for object detection*. En: Proceedings of the ICCV (2015).

[28] Ronneberger, O., Fischer, P., Brox, T., 2015. *U-net: Convolutional networks for biomedical image segmentation.* En: Medical Image Computing and Computer-Assisted Intervention.

[29] Moeskops, P., Wolterink, J. M., Velden, B. H. M., Gilhuijs, K. G. A., Leiner, T., Viergever, M. A., Isgum, I.. *Deep learning for multi-task medical image segmentation in multiple modalities.* In: Medical Image Computing and Computer-Assisted Intervention (2016).

[30] Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S.. *Deep learning based imaging data completion for improved brain disease diagnosis*. En: Medical Image Computing and Computer- Assisted Intervention (2014).

[31] Hosseini-Asl, E., Gimel'farb, G., El-Baz, A.. *Alzheimer's dis- ease diagnostics by a deeply supervised adaptable 3D convolutional network* (2016).

[32] Payan, A., Montana, G.. *Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks* (2015).

[33] Abramo, M. D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J. C., Niemeijer, M.. *Improved automated detection of di- abetic retinopathy on a publicly available dataset through integration of deep learning* (2016).

[34] Paeng, K., Hwang, S., Park, S., Kim, M., Kim, S.. *A unified framework for tumor proliferation score prediction in breast histopathology* (2016).

[35]Poudel, R. P. K., Lamata, P., Montana, G.. *Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation* (2016).

[36] Kong, B., Zhan, Y., Shin, M., Denny, T., Zhang, S.. *Recognizing end-diastole and end-systole frames via deep temporal regression network* (2016).

[37] Berkman Sahiner, Heang-Ping Chan, Nicholas Petrick, Datong Wei, Mark A. Helvie, Dorit D. Adler, and Mitchell M. Goodsitt. *Classification of Mass and Normal Breast Tissue: A Convolutional Neural Network Classifier with Spatial Domain and Texture Images* (1995).

[38] AECC. [En línea] (2014). (https://www.aecc.es/SobreElCancer/CancerPorLocalizacion/CancerMama/Paginas/cancer demama.aspx).

[39] Instituto Nacional del cáncer. [En línea] (https://www.cancer.gov/espanol/tipos/seno).

[40] World Cancer Research Fund International. Statistics. [En línea] (2012). (http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/breast-cancer-statistics).

[41] Breastcancer.org. [En línea] (2006). (http://www.breastcancer.org).

[42] Andrew R. Jamieson, Karen Drukker, Maryellen L. Giger, *Breast Image Feature Learning with Adaptive Deconvolutional Networks* (2016).

[43] Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. *Automated Mass Detection from Mammograms using Deep Learning and Random Forest*.

[44] Pablo Fonseca, Julio Mendoza, Jacques Wainer, Jose Ferrer, Joseph Pinto, Jorge Guerrero, Benjamin Castaneda. *Automatic breast density classification using a convolutional neural network architecture search procedure* (2016).

[45] Ayelet Akselrod-Ballin, Leonid Karlinsky, Sharon Alpert, Sharbell Hasoul, Rami Ben-Ari, Ella Barkan. *A Region Based Convolutional Network for Tumor Detection and Classification in Breast Mammography* (2016).

[46] John Arevalo, Fabio A. González, Raúl Ramos-Pollán, Jose L. Oliveira, Miguel Angel Guevara Lopez. *Representation learning for mammography mass lesion classification with convolutional neural networks* (2015).

[47] Dalmis, M., Gubern-Mérida, A., Vreemann, S., Karssemeijer, N., Mann, R., Platel, B., 2016. *A computer-aided diagnosis system for breast DCE-MRI at high spatiotemporal resolution*.

[48] A. Dubrovina , P. Kisilev, B. Ginsburg, S. Hashoul & R. Kimmel. *Computational mammography using deep neural networks* (2016).

[49] Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. *Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms* (2016).

[50] Sergei V. Fotin, Yin Yin, Hrishikesh Haldankar, Jeffrey W. Hoffmeister, Senthil Periaswamy. *Detection of soft tissue densities from digital breast tomosynthesis: comparison of conventional and deep learning approaches* (2016).

[51] Sangheum Hwang and Hyo-Eun Kim. *Self-Transfer Learning for Fully Weakly Supervised Object Localization* (2016).

[52] Huynh, B. Q., Li, H., Giger, M. L., Jul 2016. *Digital mammographic tumor classification using learning from deep convolutional neural networks* (2016)

[53] Michiel Kallenberg, Kersten Petersen, Mads Nielsen, Andrew Y. Ng, Pengfei Diao, Christian Igel, Celine M. Vachon, Katharina Holland, Rikke Rass Winkel, Nico Karssemeijer, and Martin Lillholm. *Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring* (2016).

[54] Kisilev, P., Sason, E., Barkan, E., Hashoul, S. *Medical image description using multi-task-loss CNN (2016).* En: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis.

[55] Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., Karssemeijer, N.. *Large scale deep learning for computer aided detection of mammographic lesions (*2016). En: Medical Image Analysis 35, 303–312.

[56] Qiu, Y., Wang, Y., Yan, S., Tan, M., Cheng, S., Liu, H., Zheng, B.. *An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology* (2016)..

[57] Samala, R. K., Chan, H.-P., Hadjiiski, L., Cha, K., Helvie, M. A.. *Deep-learning convolution neural network for computer- aided detection of*

microcalcifications in digital breast tomosynthesis (2016). En: Medical Imaging.

[58] Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Wei, J., Cha, K.. *Mass detection in digital breast tomosynthesis: Deep convolutional neural network with learning from mammography* (2016).

[59] Sun, W., Tseng, T.-L. B., Zhang, J., Qian, W.. *Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data* (2016). En: Computerized Medical Imaging and Graphics.

[60] Zhang, Q., Xiao, Y., Dai, W., Suo, J., Wang, C., Shi, J., Zheng, H.. *Deep learning based classification of breast tumors with shear-wave elastography* (2016).

[61] Kooi, T., van Ginneken, B., Karssemeijer, N., den Heeten, A.. *Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network (*2017). En: Medical Physics.

[62] Wang, J., Ding, H., Azamian, F., Zhou, B., Iribarren, C., Molloi, S., Baldi, P.. *Detecting cardiovascular disease from mammo- grams with deep learning (*2017). En: IEEE Transactions on Medical Imaging.

[63] D. Selvathi and A. Aarthy Poornila. *Breast Cancer Detection In Mammogram Images Using Deep Learning Technique* (2017).

[64] Hinton, G.E., and Salakhutdinov, R.R.. *Reducing the Dimensionality of Data with Neural Networks*. En: Science 313(5786), 504 -507 (2006).

[65] American Cancer Society (Cancer.org) [En línea] – Como entender su informe de mamograma – Puntuaje BI-RADS: (https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/como-entender-su-informe-de-mamograma.html).

[66]Mammographic Image Analysis Homepage – Databases. [En línea] (http://www.mammoimage.org/databases/)

[67] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer. *The digital database for screening mammography.*

[68] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer Jr, R. Moore, K. Chang, and S. Munishkumaran. *Current status of the digital database for screening mammography.*

[69] Gonzalez Bueno Puyal, Juana. *Desarrollo de una herramienta para la detección de tejidos anómalos en mamografías digitales mediante redes neuronales convolucionales.*

[70] Matteo Roffilli. *Advanced Machine Learning Techniques for Digital Mammography* (2016).

[71] Samir Bandyopadhyay. *Pre-processing of Mammogram Images* (2010).

[72] Aziz Makandar, Bhagirathi Halalli. *Pre-processing of Mammography Image for Early Detection of Breast Cancer* (2016).

[73] University of South Florida. Computer Vision and Patron Recognition Group – DDSM [En línea].

[74] R. K. Srivastava, K. Greff, and J. Schmidhuber. *Training very deep networks*. En:

NIPS (2015).

[75] M. Mohsin Jadoon, Qianni Zhang, Ihsan Ul Haq, Sharjeel Butt and Adeel Jadoon. *Three-Class Mammogram Classification Based on Descriptive CNN Features.*

[76]Darvin Yi, Rebecca Lynn Sawyer, David Cohn, Jared Dunnmon, Carson Lam, Xuerong Xiao, Daniel Rubin. *Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors* (2017).

[77] Carl J. Vyborny and Maryellen L. Giger. *Computer Vision and Artificial Intelligence in Mammography* (1993).

ANEXO I – ACRÓNIMOS

DL	Deep Learning
MIA	Medical Image Analysis
CNN	Convolutional Neural Networks
RNN	R Neural Networks
DNN	Deep Neural Networks
MLP	Sistema de Ayuda a la Decisión
CT	Computerized Tomography
US	Ultrasonidos
MRI	Magnetic Resonance Image
RX	Rayos-X
RBM	Restricted Boltzmann Machines
SAE	Sparse AutoEncoder
CSAE	Contractive Sparse AutoEncoder
SVM	Support Vector Machines
ROI	Region Of Interest
RBF	Radial Basis Function
NIN	Network in Network
SPP	Spatial Pyramid Pooling
FCN	Fully Connected Network
DBM	Deep Boltzman Machines
DEM	Deep Energy Models
AE	Autoencoder
DAE	Denoising Autoencoder
CAE	Contractive Autoencoder
CBIR	Content-Based Image Retrieval
CFM	Convolutional Feature Masking
HCI	Human Computer Interaction
CADe	Computer Aided Detection
CADX	Computer Aided Diagnosis
CRF	Conditional Random Field
BBDD	Bases de Datos
TUPAC	Tumor Proliferation Assessment Challenge
NLP	Natural Language Processing